REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)* 31/Jul/2001	2. REPORT TYPE DISSERTATION	3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
AN ARTIFICIAL NEURAL NETWORK MODEL FOR THE
PREDICTION OF CHILD PHYSICAL ABUSE RECURRENCE

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
CAPT FLAHERTY CHRISTOPHER W

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF TENNESSEE KNOXVILLE

8. PERFORMING ORGANIZATION REPORT NUMBER
CI01-215

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

13. SUPPLEMENTARY NOTES

14. ABSTRACT

20010904 051

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		165	19b. TELEPHONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

An Artificial Neural Network Model for the

Prediction of Child Physical Abuse Recurrences

Christopher W. Flaherty, Captain, U.S. Air Force

2001

165 pages

Ph.D. in Social Work

The University of Tennessee

Abstract

The present study explored the potential of an artificial neural network to improve prediction of recurrences of child physical abuse. Conducted on electronic data file compiled by the U.S. Air Force's central registry of child abuse reports, selected variables pertaining to all child physical abuse reports received from 1990-2000 (N=5612) were examined. Thirteen predictor variables and five interaction terms were identified for analysis.

It was hypothesized that each of the thirteen predictor variables and five interaction terms would be correlated with abuse recurrence when controlling for all other variables in the model. Using binary logistic regression (BLR) to analyze data, only four of the main effect variables and one interaction term were correlated with abuse recurrence.

It was also hypothesized that an artificial neural network model would predict abuse recurrences better than an alternative method (BLR) due to superior ability to model complex interactions and curvilinear relationships among the selected variables. The hypothesis was not confirmed. Although both methods predicted recurrences significantly better than chance, the BLR model produced predictions that were slightly, but significantly better than the ANN Model. The BLR was also advantageous in terms of providing more information regarding contributions of individual predictors.

It was concluded that both BLR and ANNs offer powerful tools to be used in future efforts to build abuse prediction models. When applied to the present data, BLR was more useful.

BIBLIOGRAPHY

Abramczyk, L.W., & Swiegart, C. (1985). Child abuse and neglect: Indicated versus unfounded report characteristics. Final report to DHHS/OHDS/NCAAN for Grant Award #90-CA-0944.

Adams, W., Barone, N., & Tooman, P. (1982). The dilemma of anonymous reporting in child protective services. Child Welfare, 61, 3-14.

Air Force ServeNet, Family Advocacy Program. (2000). Available: http://airforcefap.org/fap/menu.asp

American Humane Association. (1988). Highlights of official child neglect and abuse reporting, 1986. Denver, CO: Author.

Ammerman, R.T., Kolko, D.J., Kirisci, L., Blackson, T.C., & Dawes, M.A. (1999). Child abuse potential in parents with histories of substance abuse disorder. Child Abuse & Neglect, 23(12), 1225-1238.

Anderson, R., Ambrosino, R., Valentine, D., & Lauderdale, M. (1983). Child deaths attributed to abuse and neglect: An empirical study. Children and Youth Services Review, 5, 75-89.

Ashizawa, K., MacMahon, H., Ishida, T., Nakamura, K., Vyborny, C.J., Katsuragawa, S., & Doi, K. (1999). Effect of an artificial neural network on radiologists' performance in the differential diagnosis of interstitial lung disease using chest radiographs. American Journal of Roentgenology, 172(5), 1311-5.

Baird, C. (1988). Development of risk assessment indices for the Alaska Department of Health and Social Services. In T. Tatara (Ed), Validation research in CPS

risk assessment: Three recent studies (pp. 85-121). Washington, DC: American Public Welfare Association.

Baird, C., Ereth, J., & Wagner, D. (1999). Research-based risk assessment: Adding equity to CPS decision-making [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Baird, C., Wagner, D., Caskey, R., & Neuenfeldt, D. (1995). The Michigan Department of Social Services Structured Decision Making System: An evaluation of its impact on child protection services [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Baird, C., Wagner, D., Healy, & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. Child Welfare, 78(6), 723-748.

Baumann, D.J., Esterline, J.A., Zuniga, G., Smith, S., Whiteside, D., Fluke, J., Goertz, B., & Cohen, M. (1997). The implementation of risk assessment. In Worker Improvement the the Structured Decision and Outcome Model. Texas Department of Protective and Regulatory Services.

Baxt, W.G., & Skora, J. (1996). Prospective validation of artificial neural network trained to identify myocardial infarction. The Lancet, 347, 12-15.

Bergman, A.B., Larsen, R.M., & Mueller, B.A. (1986). Changing spectrum of serious child abuse. Pediatrics, 77, 113-116.

Berkowitz, S. (1991). Key findings on definitions of risk assessment to children and uses of risk assessment by state CPS agencies from the state survey component of the

Study of High Risk Child Abuse and Neglect groups. Paper presented at the National Center on Child Abuse and Neglect. Washington, DC.

Bolen, R.M. (1998). Predicting risk to be sexually abused: A comparison of logistic regression to event history analysis. Child Maltreatment, 3(2), 157-170.

Brissett-Chapman, S. (1997). Child protection risk assessment and African American children: Cultural ramifications for families and communities. Child Welfare, 74(1), 45-63.

Brodzinski, J.D., Crable, E.A., & Sherer, R.F. (1994). Using artificial intelligence to model juvenile recidivism patterns. Computers in Human Services, 10(4), 1-19.

Bryce, T.J., Dewhirst, M.W., Floyd, C.E., Hars, V., & Brizel, D.M. (1998). Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck. International Journal of Radiation, Oncology, Biology, and Physics, 41(2), 339-45.

Buchanan, B.G., & Shortliffe, E.H. (Eds.). (1984). Rule-based expert systems. Reading, MA: Addison-Wesley.

Burrell, B., Thompson, B., & Sexton, D. (1994). Predicting child abuse potential across family types. Child Abuse & Neglect, 18(12), 1039-1049.

Calica, R., Cotton, E., & Edwards, M. (1998). Illinois child endangerment risk assessment protocol: Impact on short-term recurrence rates-year two. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Camasso, M.J., & Jagannathan, R. (1995). Prediction accuracy of the Washington and Illinois risk assessment instruments: An application of receiver operating characteristic curve analysis. Social Work Research, 19(3), 174-183.

Campana, A., Duci, A., Gambini, O., & Scarone, S. (1999). An artificial neural network that uses eye-tracking performance to identify patients with schizophrenia. Schizophrenia Bulletin, 25(4), 789-799.

Chaffin, M., Kelleher, & Hollenberg, J. (1996). Onset of physical abuse and neglect: Psychiatric, substance abuse, and social risk factors from prospective community data. Child Abuse & Neglect, 3, 191-203.

Child Abuse Prevention and Treatment Act (CAPTA) (P.L. 93-247, Jan. 1996 version), 42 U.S.C. 5101. Originally passed in 1974.

Church, K.B., & Curram, S.P. (1996). Forecasting consumers' expenditure: A comparison between econometric and neural network models. International Journal of Forecasting, 12(2), 255-67.

Cohen, J. (1977). Statistical power analyses for the behavioral sciences. New York: Academic Press.

Collier, A.F., McClure, F.H., Collier, J., Otto, C., & Polloi, A. (1999). Culture-Specific views of child maltreatment and parenting styles in a Pacific-Island community. Child Abuse & Neglect, 23(3), 229-244.

Collins, J.M., & Clark, M.R. (1993). An application of the theory of neural computation to the prediction of workplace behavior: an illustration and assessment of network analysis. Personal Psychology, 46, 503-524.

Connelly, C.D., & Straus, M.A. (1992). Mother's age and risk for physical abuse. Child Abuse & Neglect, 16, 709-718.

Cooil, B., Winer, R.S., & Rados, D.L. (1987). Cross-validation for prediction. Journal of Marketing Research, 24, 271-279.

Creighton, S.J. (1985). Am epidemiological study of abused children and their families in the United Kingdom between 1977 and 1982. Child Abuse and Neglect, 9, 441-448.

Cross, S.S., Harrison, R.F., & Kennedy, R.L. (1995). Introduction to neural networks. The Lancet, 346, October 21, 1075-1079.

Daley, M.R., & Pilavin, I. (1982). "Violence against children" revisited: Some necessary clarification of findings from a major national study. Journal of Social Service Research, 5(1/2), 61-81.

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgement. Science, 243, 1668-1674.

DePanfilis, D., & Scannapieco, M. (1994). Assessing the safety of children at risk of maltreatment: Decision-making models. Child Welfare, 73(3), 229-245.

DePanfilis, D., & Zuravin, S.J. (1999). Epidemiology of child maltreatment recurrences. Social Service Review, 73(2), 218-239.

DePanfilis, D., & Zuravin, S.J. (1999). Predicting child maltreatment recurrences during treatment. Child Abuse & Neglect, 23(8), 729-743.

Department of Defense. (1986). Family Advocacy Program (Directive 6400.1) Washington, D.C.: Author.

Department of Defense. (1987). Child and Spouse Abuse Report (Directive 6400.2). Washington, D.C.: Author.

Department of Defense. (1992). Family Advocacy Program (Directive 6400.1) Washington, D.C.: Author.

Department of Health and Human Services, National Center on Child Abuse and Neglect. (1995). Child maltreatment 1993: Reports from the states to the Center on Child Abuse and Neglect. Washington, DC: U.S. Government Printing Office.

de Tommaso, M., Sciruicchio, V., Bellotti, R., Castellano, M., Tota, P., Guido, M., Sasanelli, G., & Puca, F. (1997). Discrimination between migraine patients and normal subjects based on steady state visual evoked potentials: Discriminant analysis and artificial neural network classifiers. Functional Neurology, 12(6), 333-338.

Doueck, H.J., English, D.J., DePanfilis, D., & Moote, G.T., jr. (1993). Decision making in child protective services: A comparison of selected risk assessment systems. Child Welfare, 72, 441-453.

Doueck, H.J., Levine, M., & Bronson, D.E. (1993). Risk assessment in child protective services: An evaluation of the Child at Risk Field System. Journal of Interpersonal Violence, 8(4), 446-467.

Doueck, H.J., & Lyons, P. (1998). The Child Well-Being Scales as a predictor of caseworker attention and services in child protection: A preliminary analysis. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Downing, J.D., Wells, S.J., & Fluke, J. (1990). Gatekeeping in child protective services: A survey of screening policies. Child Welfare, 69(4), 357-369.

Drake, B. (1996). Unraveling "unsubstantiated". Child Maltreatment, 1(5), 261-271.

Dwyer, M. (1992). A comparison of statistical techniques and artificial neural network models in corporate bankruptcy prediction. Doctoral dissertation, University of Wisconcin-Madison.

Dybowski, R., Weller, P., Chang, R., & Grant, V. (1996). Prediction of outcome in critically ill patients using artificial neural network synthesized by genetic algorithm. The Lancet, 347(9009), 1146-51.

El-Solh, A.A., Hsiao, C., Goodnugh, S., Serghani, J., & Grant, B.J.B. (1999). Predicting active pulmonary tuberculosis using an artificial neural network. Chest, 116(4), 968-973.

English, D.J. (1996). The promise and reality of risk assessment. Protecting Childern, 12(2), 9-13.

English, D.J., Marshall, D.B., Coghlan, L., Brummel, S., & Orme, M. (2000). Causes and consequences of the substantiation decision in Washington state child protective services. Children & Youth Services Review, to appear.

English, D.J., & Pecora, P.J. (1994). Risk assessment as a practice method in child protective services. Child Welfare, 73(5), 451-468.

Everson, H.T., er al. (1994). Using artificial neural networks in educational research: some comparisons with linear statistical models. Paper presented a the Annual Meeting of the Council on Measurement in Education, New Orleans, LA, 5-7 April.

Fanshel, D., Finch, S.J., & Grundy, J.F. (1994). Testing the measurement properties of risk assessment instruments in child protective services. Child Abuse & Neglect, 18, 1073-1084.

Farmer, R.M., Medearis, A.L., Hirata, G.I., & Platt, L.D. (1992). The use of a neural network for the ultrasonographic estimation of fetal weight in the macroscopic fetus. American Journal of Obstetrics and Gynecology, 166(5), 1467-1472.

Ferrara, J.M., Parry, J.D., & Lubke, M.M. (1985). CLASS.1: An expert system for student classification. Technical paper (ERIC Document Reproduction Service No. ED 263-734), Utah State University, Logan, UT.

Flango, V.E. (1991). Can central registries improve substantiation rates in child abuse and neglect cases? Child Abuse & Neglect, 15, 403-413.

Fluke, J. (1993). The use of expert systems and neural networks in CPS risk assessment: What are the possibilities and limitations. Summary of small group discussion at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J. (1994). Emerging critical conceptual issues in risk assessment research and practice. Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J., Wells, S., England, P., & Gamble, T. (1993). Evaluation of the Pennsylvania approach to risk assessment. Paper presented at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J., Edwards, M., & Wells, S. (1996). Evaluation of the Illinois child endangerment risk assessment protocol: Safety domain. Unpublished manuscript.

Fluke, J.D., Yuan, Y.T., & Edwards, M. (1999). Recurrence of maltreatment: An application of the National Child Abuse and Neglect Data System (NCANDS). Child Abuse & Neglect, 23(7), 633-650.

Fryer, G.E., & Miyoshi (1994). A survival analysis of the revictimization of children: The case of Colorado. Child Abuse & Neglect, 18(12), 1063-1071.

Fuller, T.L., & Wells, S.J. (1998). Illinois child endangerment risk assessment protocol: A case control study of short-term recurrence among intact family cases. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Galletly, C.A., Clark, C.R., & McFarlane, A.C. (1996). Artificial neural networks: A prospective tool for the analysis of psychiatric disorders. Journal of Psychiatry and Neuroscience, 21(4), 239-247.

Gallinari, P., Thiria, S., Badran, F., & Fogelman-Soulie, F. (1991). On the relations between discriminant analysis and multilayer perceptrons. Neural Networks, 4, 349-360.

Garson, G.D. (1998). Neural Networks: An Introductory Guide for Social Scientists. Thousand Oaks, California. Sage.

Gaudin, J.M., Jr., Polansky, N., & Kilpatrick, A.C. (1992). The Child Well-Being Scales: A field trial. Child Welfare, 17, 319-328.

Gelles, R.J. (1982). Child abuse and family violence: Implications for medical professionals. In Newberger, E.H. (Ed), Child Abuse. Boston. Little Brown.

Gelles, R.J. (1989). Child abuse and violence in single-parent families: Parent absence and economic deprivation. American Journal of Orthopsychiatry, 59(4), 492-501.

Gil, D.G. (1970). Violence against children. Cambridge, MA: Harvard University Press.

Gingerich, W.J. (1990). Expert systems and their potential uses in social work. Families in Society, 71(4), 220-228.

Giovannoni, J.M., & Becerra, R.M. (1979). Defining child abuse. New York. Free Press.

Goodman, H., Gingerich, W.J., & de Shazer, S. (1989). BRIEFER: An expert system for clinical practice. Computers in Human Services, 5(1/2), 53-68.

Gordon, J.S. (1991). (Probability correct classification as a function of increasing decision thresholds). Unpublished raw data.

Gordon, J.S. (1992). A neural network approach to the prediction of violence. Unpublished doctoral dissertation. Oklahoma State University.

Greenland, C. (1987). Preventing CAN deaths: An international study of deaths due to child abuse and neglect. London: Tavistock Publications.

Hampton, R.L. (1987). Race, class, and child maltreatment. Journal of Comparative Family Studies, 18, 113-126.

Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143, 29-36.

Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating curves derived from the same cases. Radiology, 148, 839-843.

Hartzberg, J., Stanley, J., & Lawrence, M. (1990). Brainmaker user's guide and reference manual (Computer program manual). Sierra Madre, CA: California Scientific Software.

Heckert, T.M. (1994). Effect of base rate on the applicability of neural network analysis for classification. Doctoral dissertation, Bowling Green State University.

Hedlund, J.L., Viewing, B.W., & Cho, D.W. (1987). Computer consultation for emotional crises: An expert system for "non-experts." Computers in Human Behavior, 3, 109-127.

Herrenkohl, R.C., Herrenkohl, E.C., Egolf, B., & Seech, M. (1979). The repetition of child abuse: How frequently does it occur? Child Abuse & Neglect, 3, 67-72.

Hiemstra, Y. (1996). Linear regression versus backpropagation networks to predict quarterly stock market excess returns. Computational Economics, 9(1), 67-76.

Holder, W., & Corey, M.K. (1993). The Child At Risk Field System: Forms and instructions booklet. Charlotte, NC: Action for Child Protection.

Howing, P.T., Wodarski, J.S., Kurtz, P.D., & Gaudin, J.M., Jr. (1989). Methodological issues in child maltreatment research. Social Work Research and Abstracts, 25(3), 3-7.

Hutchinson, E.D. (1989). Child protective screening decisions: An analysis of predictive factors. Social Work Research and Abstracts, 25(3), 9-15.

Inkelas, M, & Halfon, N. (1997). Recidivism in child protective services. Children and Youth Services Review, 19, 139-161.

Jackson, A. (1983). Professional and non-professional sources of child maltreatment reports. For Your Information, 4, 3-7.

Jason, J., & Andereck, N.D. (1983). Fatal child abuse in Georgia: The epidemiology of severe physical child abuse. Child Abuse & Neglect, 7, 1-9.

Jefferson, M.F., Pendleton, N., Lucas, C.P., Lucas, S.B., & Horan, M.A. (1998). Evolution of artificial neural network architecture: Prediction of depression after mania. Methods of Information in Medicane, 37, 220-225.

Johnson, W. (1993). <u>Maltreatment recurrence as a criterion for validating risk assessment instruments.</u> Paper presented at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Johnson, W. (1994). <u>Culturally sensitive risk assessment: A conceptual framework and supporting data.</u> Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Johnson, W. (1996). Risk assessment research: progress and future directions. <u>Protecting Children, 12</u>(2), 14-19.

Johnson, W., & Clancy, T. (1988). A study to find improved methods of screening and disposing of reports of child maltreatment in the emergency response program in Alameda County, California. In Tatara (Ed), <u>Validation research in CPS risk assessment: Three recent studies.</u> Occasional Monograph Series, No. 2 (pp47-84). Washington, DC: American Public Welfare Association.

Johnson, W., & Clancy, T. (1989). <u>Preliminary findings from a study of risk assessment accuracy and service effectiveness in home-based services physical abuse cases.</u> In T. Tatara (Ed.), Third National Roundtable on CPS Risk Assessment: Summary of highlights. Washington, DC: American Public Welfare Association.

Johnson, W., & L'Esperance, J. (1984). Predicting the recurrence of child abuse. <u>Social Work Research and Abstracts, 20</u>(2), 21-26.

Kennedy, R.L., Harrison, R.F., Burton, A.M., Fraser, H.S., Hamer, W.G., MacArthur, D., McAllum, R., Steedman, D.J. (1997). An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency

department: evaluation and comparison with serum myoglobin measurements. Computer Methods and Programs in Biomedicine, 52, 93-103.

Kolko, D.J., (1998). CPS operations and risk assessment in child abuse cases receiving services: Initial findings from the Pittsburgh service delivery study. Child Maltreatment, 3(3), 262-275.

Korbin, J.E. (1981). Very few cases of child abuse and neglect in the People's Republic of China. In J.E. Korbin (Ed.), Child abuse: Cross-cultural perspectives. Berkeley, CA: University of California Press.

Kruttschnitt, C., Mcleod, J.D., & Dornfeld, M. (1994). The economic environment of child abuse. Social Problems, 41(2), 299-315.

Law, J.R., Maumann, D.J., Gober, K.J., Schultz, D.F., Ohmart, R., & Kern, H. (1997). Phase III: Evaluating the effectiveness of an actuarial risk assessment model. In Worker Improvement the the Structured Decision and Outcome Model. Texas Department of Protective and Regulatory Services.

Law, J.R., Kern, H.D., Schultz, D.F., Gober, K.J., Schwab, J., & Baumann, D.J. (1997). Screening models for intake and investigation. In Worker Improvement the the Structured Decision and Outcome Model. Texas Department of Protective and Regulatory Services.

Law, R., McFadden, T., & Kern, H. (1994). Risk assessment at intake and at investigation. Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Lawrence, J. (1993). Introduction to neural networks: Design, theory and application (5th ed.) Nevada City, CA: California Scientific Software.

Levy, H.B., Markovic, J., Chaudhry, U., Ahart, S., & Torres, H. (1995). Reabuse rates in a sample of children followed for 5 years after discharge from a child abuse inpatient assessment program. Child Abuse & Neglect, 19, 1363-1377.

Lindsey, D. (1991). Factors affecting the foster care placement decision: Analysis of national survey data. American Journal of Orthopsychiatry, 61, 272-281.

Lindsey, D. (1992). Reliability of the foster care placement decision: A review. Research on Social Work Practice, 2, 65-80.

Lippman, R.P. (1987). An introduction to computing with neural nets. IEEE ASSP Magazine, April, pp. 4-22.

Luttrell, J., Hull, S., & Wagner, D. (1995). The Michigan Department of Social Services structured decision-making system. Paper presented as the Ninth National Roundtable on CPS Risk Assessment. San Francisco, California.

Lutzker, J.R., & Rice, J.M. (1987). Using recidivism data to evaluate Project 12-Ways: An ecobehavioral approach to the treatment and prevention of child abuse and neglect. Journal of Family Violence, 2, 283-290.

Lyons, P., Doueck, H.J., Koster, A.J., Witzky, M.K., & Kelly, P.L. (1999). The Child Well-Being Scales as a clinical tool and a management information system. Child Welfare, 78(2), 241-258.

Lyons, P., Doueck, H.J., & Wodarski, J.S. (1996). Risk assessment for child protective services: A review of the empirical literature on instrument performance. Social Work Research, 20(3), 143-155.

Magura, S., Moses, B.S., & Jones, M.A. (1987). Assessing risk and measuring

change in families: The Family Risk Scales. Washington, DC: Child Welfare League of

America.

March, J.S., & Curry, J.F. (1998). Predicting the outcome of treatment. Journal of

Abnormal Child Psychology, 26(1), 39-51.

Marks, J., & McDonald, T.P. (1989). Risk assessment in child protective services:

Predicting recurrence of child maltreatment. Portland, ME: National Child Welfare

Resource Center.

Marshall, D.B., & English, D.J. (2000). Neural network modeling of risk

assessment in child protective services. Psychological Methods, 5(1) in press.

Martindale, E.S., Ferrara, J.M., & Campbell, B.W. (1987). A preliminary report

on the performance of CLASS.LD. Computers in Human Behavior, 3, 263-272.

McDonald, T., & Marks, J. (1991). A review of risk factors assessed in child

protective services. Social Service Review, 65(1), 112-132.

Meehl, P.E. (1954). Clinical versus statistical prediction. Minneapolis, Minnesota.

University of Minnesota Press.

Meyer, B.L. (1998). Implementing actuarial risk assessment: Policy decisions and

field practice in New Mexico. Paper presented at the Twelfth National Roundtable on

CPS Risk Assessment. San Francisco, California.

Meyer, B.L., & Wagner, D. (1998). Using actuarial risk assessment to identify

unsubstantiated cases for preventative intervention in New Mexico. Paper presented at

the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Milner, J.S., & Wimberley, R.C. (1979). An inventory for the identification of child abusers. Journal of Clinical Psychology, 35, 95-100.

Milner, J.S. (1995). Physical child abuse assessment: Perpetrator evaluation. In J.C. Campbell (Ed.), Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 41-67). Thousand Oaks, California. Sage.

Milner, J.S., & Campbell, J.C. (1995). Prediction issues for practitioners. In J.C. Campbell (Ed.), Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 20-40). Thousand Oaks, California. Sage.

Mollerstrom, W.W., Patchner, M.A., & Milner, J.S. (1995). Child maltreatment: the United States Air Force's response. Child Abuse & Neglect, 19(3), 325-

Mulsant, B., & Servan-Schreiber, D. (1984). Knowledge engineering: A daily activity on a hospital ward. Computers in Biomedical Research, 17, 71-91.

Munro, E. (1999). Common errors of reasoning in child protection work. Child Abuse & Neglect, 23(8). 745-758.

Murphy-Berman, V. (1994). A conceptual framework for thinking about risk assessment and case management in child protective service. Child Abuse & Neglect, 18(2), 193-201.

National Center for Child Abuse and Neglect (1981). National study of the incidence and severity of child abuse and neglect. 81-30325. Washington, DC: U.S. Department of Health and Human Services. [NIS-1].

National Center for Child Abuse and Neglect (1988). National study of the incidence and severity of child abuse and neglect. Washington, DC: U.S. Department of Health and Human Services. [NIS- 2].

National Center for Child Abuse and Neglect (1996). <u>National study of the incidence and severity of child abuse and neglect.</u> 105-91-1800. Washington, DC: U.S. Department of Health and Human Services. [NIS- 3].

National Council on Crime and Delinquency (NCCD), (2000). <u>Child abuse and neglect: Improving consistency in decision-making</u> [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Nasuti, J.P., & Pecora, P.J. (1993). Risk assessment scales in child protection: A test of the internal consistency and interrater reliability of one statewide system. <u>Social Work Research and Abstracts, 29</u>(2), 28-33.

Neuenfeldt, D. & DeMares, M. (1994). <u>Risk assessment validation study and its use in four urban Wisconsin counties.</u> Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Norusis, M.J. (1997). <u>SPSS Guide to Data Analysis.</u> Upper Saddle River, NJ: Prentice-Hall, Inc.

Nunnally, J.C. (1967). <u>Psychometric Theory.</u> New York, New York. McGraw-Hill.

Ongphiphadhanakul, B., Rajatanavin, R., Chailurkit, L., Piaseu, N., Teerarungsikul, K., Sirisriro, R., Komindr, S., & Pauvilai, G. (1997). Prediction of low bone mineral density in post menopausal women by artificial neural network model compared to logistic regression model. <u>Journal of the Medical Association of Thailand, 80</u>(8), 508-515.

Orme, J.G. (1986). Erroneous use of observed percentage agreement in discriminant function analysis. <u>Social Work Research and Abstracts, 22,</u> 2.

Paddock, J.B. (1995). Factors predictive of injury severity in cases of physical child abuse among Air Force families: A cross-validation study. Unpublished dissertation.

Parton, N., Thorpe, D., & Wattam, C. (1997). Child Protection: Risk and the Moral Order. London. Macmillan Press.

Patterson, D.A. & Cloud, R.N. (2000). The application of artificial neural networks for outcome prediction in a cohort of severely mentally ill outpatients. Forthcoming in The Journal of Technology for Human Services.

Pecora, P.J. (1991). Investigating allegations of child maltreatment: The strengths and limitations of current risk assessment systems. Child and Youth Services, 15(2), 73-92.

Picard, R.R., & Cook, R.D. (1984). Cross-validation of regression models. Journal of the American Statistical Association, 79(387), 575-583.

Pofahl, W.E., Walczak, S.M., Rhone, E., & Izenberg, S. (1998). Use of an artificial neural network to predict length of stay in acute pancreatitis. The American Surgeon, 64(9), 868-872.

Pugh, G.A. (1991). A comparison of neural networks to SPC charts. Computers and Industrial Engineering, 21(1), 253-55.

Ross, S.M. (1996). Risk of physical abuse to children of spouse abusing parents. Child Abuse & Neglect, 20(7), 589-598.

Rosenthal, J.A. (1988). Patterns of reported child abuse and neglect. Child Abuse & Neglect, 12, 263-271.

Ross, S.M. (1996). Risk of physical abuse to children of spouse abusing parents. Child Abuse & Neglect, 20(7), 589-598.

Rossi, P., Schuerman, & Budde, S. (1996). Understanding child maltreatment decisions and those who make them. Chicago: University of Chicago, Chapin Hall Center for Children.

Rubin, A., & Babbie, E. (1997). Research methods for social work (3rd Ed.), Pacific Grove, CA: Brooks/Cole.

Ruscio, J. (1998). Information integration in child welfare cases: An introduction to statistical decision-making. Child Maltreatment, 3(2), 143-156.

Salovitz, B. (1992). New York State risk assessment and service planning model: A review of the developmental process and key features of the model. Paper presented at the Sixth National Roundtable on CPS Risk Assessment. San Francisco, California. American Public Welfare Association.

Schene, P. (1996). The risk assessment roundtables: a ten-year perspective. Protecting Children, 12(2), 4-8.

Schoech, D. (1999). Human service technology: Understanding, designing, and implementing computer and internet applications in the social services. New York, New York. The Haworth Press, Inc.

Schuerman, J.R. (1987). Expert consulting systems in social welfare. Social Work Research and Abstracts, 23(3), 14-18.

Schuerman, J.R., & Vogel, L.H. (1986). Computer support of placement planning: The use of expert systems in child welfare. Child Welfare, 65(6), 531-543.

Seaberg, J.R. (1977). Predictors of injury in physical child abuse. Journal of Social Service Research, 1(1), 63-76.

Sedlak, A.J. (1997). Risk factors for the occurrence of child abuse and neglect. Journal of Aggression, Maltreatment, and Trauma, 1(1), 149-187.

Sedlack, A.J., & Broadhurst, D.B. (1996). Executive summary of the third national incidence study of child abuse and neglect. U.S. Department of Health and Human Services 105-91-1800.

Serna, R.W., Baer, R.D., & Ferrara, J.M. (1986). An expert system for behavior analysis. Paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Shadish, W. H., Jr. (1986). Planned critical multiplism: Some elaborations. Behavioral Assessment, 8, 75-103.

Shapiro, D. (1979). Parents and protectors: A study of child abuse and neglect. New York Research Center, Child Welfare League of America.

Sheets, D.A. (1996). Caseworkers, computers and risk assessment: A promising partnership. APSAC Advisor, 9(1), 7-12.

Sohl, J.-H.E. & Venkatachalalm, A.R. (1995). A neural network approach to forcasting-model selection, Information Management, 29(6), 297-303.

SPSS 10.0 (1999). [Computer program]. Chicago: SPSS Inc.

SPSS (1999) SPSS 10.0 User's Guide, Chicago: Author.

SPSS Neural Connection 2.0 (1997). [Computer program]. Chicago: SPSS Inc.

SPSS (1997). Neural Connection 2.0 User's Guide, Chicago: Author.

Squadrito, E. & Wagner, D. (1993). Family risk assessment study by the Rhode Island Department of Children, Youth, and Families and the National Council on Crime and Delinquency/Children's Research Center. Paper presented at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Starr, R.H. (Ed.). (1982). Child abuse prediction: Policy implications. Cambridge, Massachusetts: Ballinger.

Tailor, A., Jurkovic, D., Bourne, T.H., Collins, W.P., & Campbell, S. (1999). Sonographic prediction of malignancy in adnexal masses using an artificial neural network. British Journal of Obstetrics and Gynaecology, 106, 21-30.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62(2), 157-176.

Timmerman, D., Verrelst, H., Bourne, T.H., De Moor, B., Collins, W.P., Vergote, I., & Vandewalle, J. (1999). Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound in Obstetrics and Gynecology, 13(1), 17-25.

U.S. Department of Health and Human Services. Child maltreatment (1997): Reports from the states to the National Child Abuse and Neglect Data System (Washington, DC: U.S. Government Printing Office.

Valentine, D.P., Acuff, D.S., Freeman, M.L., & Andreas, T. (1984). Defining child maltreatment: A multidisciplinary overview. Child Welfare, 63(6), 497-509.

Vogt, W.P. (1999). Dictionary of statistics and methodology: A nontechnical guide for the social sciences. Thousand Oaks, CA: Sage.

Wald, M.S., & Woolverton, M. (1990). Risk assessment: "The emperor's new clothes?" Child Welfare, 69, 483-511.

Weedon, J., Torti, T.W., & Zunder, P. (1988). Vermont Division of Social Services Family Risk Assessment Matrix: Research and Evaluation. In Tatara (Ed), Validation research in CPS risk assessment: Three recent studies (pp. 3-43). Washington, D.C.: American Public Welfare Association.

Weiss, S.M., & Kurlikowski, C.A. (1991). Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems. San Mateo, CA: Morgan Kaufmann.

Wells, S.J., & Anderson, T. (1992). Workers' estimation of risk as a predictor of case substantiation. Washington, DC: American Public Welfare Association.

Wells, S.J., & Anderson, T. (1992). Model building in child protective services intake and investigation. Final report to the National Center on Child Abuse and Neglect for Grant #90-CA-1407. Washington, DC: American Bar Association Center on Children and the Law.

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward neural models. Journal of the American Statistical Society, 84, 1003-1013.

Wolfner, G.D., & Gelles, R.J. (1993). A profile of violence toward children: A national study. Child Abuse & Neglect, 17, 197-212.

Zaknich, A. (1998). Artificial neural networks: An introductory course [On-line]. Available: http://www.maths.uwa.edu.au/~rkcalley/ann_all/ann_all.html

Zernikow, B., Holtmannspoetter, K., Michel, E., Theilhaber, M., Pielemeier, W., & HenneckeK.H. (1998). Artificial neural network for predicting intercranial haemorrhage in preterm neonates. Acta Paediatrica, 87(9), 969-75.

Zernikow, B., Holtmannspoetter, K., Michel, E., Pielemeier, W., Hornschuh, F., Westermann, A., & Hennecke, K.H. (1998). Artificial neural network for risk assessment in preterm neonates. Archives of Disease in Childhood, 79(2), F129-34.

Zhang, Z., Barnhill, S.D., Zhang, H., Xu, F., Yu, Y., Jacobs, I., Woolas, R.P., Berchuck, A., Madyastha, K.R., & Bast, R.C, jr. (1999). Combination of multiple serum markers using an artificuial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecologic Oncology, 73, 56-61.

Zhao, R., Xu, G., Yue, B., Liebich, H.M., & Zhang, Y. (1998). Artificial neural network classification based on capillary electrophoresis of urinary nucleosides for the clinical diagnosis of tumors. Journal of Chromatography A, 828, 489-496.

Zou, Y., Shen, Y., Shu, L., Wang, Y., Feng, F., Xu, K., Qu, Y., Song, Y., Zhong, Y., Wang, M., & Liu, W. (1996). Artificial neural network to assist psychiatric diagnosis. British Journal of Psychiatry, 169(1), 64-67.

Zuravin, S. (1988). Child maltreatment and teenage first birth: A relationship mediated by chronic sociodemographic stress? American Journal of Orthopsychiatry, 58, 91-103.

Zuravin, S.J., Orme, J.G., & Hegar, R.L. (1991). Factors predicting substantiation and severity of child abuse injury. In National Center on Child Abuse and Neglect: Symposium on Risk Assessment in Child Protective Services (Vol. 4). Washington, DC: National Center on Child Abuse and Neglect.

Zuravin, S.J., Orme, J.G., Hegar, R.L. (1994). Predicting severity of child abuse injury with ordinal probit regression. <u>Social Work Research, 18</u>(3), 131-138.

Zuravin, S.J., Orme, J.G., Hegar, R.L. (1995). Disposition of child physical abuse reports: Review of the literature and test of a predictive model. <u>Children and Youth Services Review, 17</u>(4), 547-566.

Zuravin, S.J., Watson, B., & Ehrenschaft, M. (1987). Anonymous reports of child physical abuse: Are they as serious as reports from other sources? <u>Child Abuse & Neglect, 11,</u> 521-529.

An Artificial Neural Network Model for the Prediction
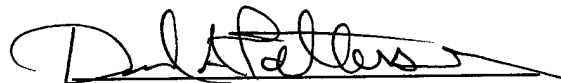
of Child Physical Abuse Recurrences

A Dissertation

Presented for the

Doctor of Philosophy

Degree

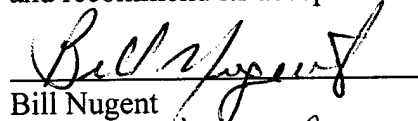The University of Tennessee, Knoxville
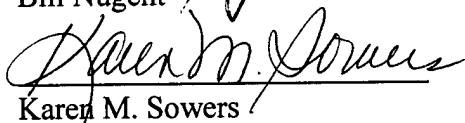
Christopher W. Flaherty

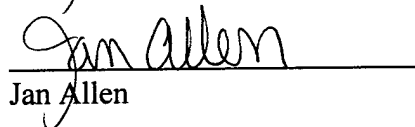August 2001

To the Graduate Council:

I am submitting herewith a dissertation written by Chris Flaherty entitled "An Artificial Neural Network Model for the Prediction of Child Physical Abuse Recurrences." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Social Work.

David A. Patterson, Major Professor

We have read this dissertation
and recommend its acceptance:

Bill Nugent

Karen M. Sowers

Jan Allen

Accepted for the Council:

Interim Vice Provost and
Dean of the Graduate School

Abstract

All 50 states have passed some form of mandatory reporting law to qualify for

funding under the Child Abuse Prevention and Treatment act of 1974 (P.L. 93-247).

Consequently, child protective service (CPS) agencies have experienced a dramatic

increase in reports of abuse and neglect without corresponding increases in funding over

the past several years. In response, many CPS agencies have turned to formal risk

assessment systems to aid caseworker in making various decisions. Various

methodological obstacles have impeded efforts to predict child abuse.

The present study explored the potential of an artificial neural network to improve

prediction of recurrences of child physical abuse. Conducted on electronic data file

compiled by the U.S. Air Force's central registry of child abuse reports, selected

variables pertaining to all child physical abuse reports received from 1990-2000

(N=5612) were examined. Thirteen predictor variables and five interaction terms were

identified for analysis.

It was hypothesized that each of the thirteen predictor variables and five

interaction terms would be correlated with abuse recurrence when controlling for all other

variables in the model. Using binary logistic regression (BLR) to analyze data, only four

of the main effect variables and one interaction term were correlated with abuse

recurrence.

It was also hypothesized that an artificial neural network model would predict

abuse recurrences better than an alternative method (BLR) due to superior ability to

model complex interactions and curvilinear relationships among the selected variables.

The hypothesis was not confirmed. Although both methods predicted recurrences significantly better than chance, the BLR model produced predictions that were slightly, but significantly better than the ANN Model. The BLR was also advantageous in terms of providing more information regarding contributions of individual predictors.

It was concluded that both BLR and ANNs offer powerful tools to be used in future efforts to build abuse prediction models. When applied to the present data, BLR was more useful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Statement of the Problem

The number of official reports of child abuse and neglect has consistently and dramatically increased over the past 25 years. All 50 states have passed some form of mandatory child abuse and neglect reporting law in order to qualify for funding under the Child Abuse Prevention and Treatment Act (CAPTA) of 1974 (P.L. 93-247). Consequently, official incidence studies have shown an increase in reports of child maltreatment received by Child Protective Services (CPS) agencies from 416,033 reports in 1976 to 1,700,000 reports in 1990 (NCCAN 1981, 1988, 1992) and in 1997, 3 million children were the alleged victims of maltreatment in the 45 states that reported these data (U.S. Department of Health and Human Services, 1997). Although neglect accounts for the highest percentage of reports (56%)(DHHS, 1997), the estimated number of physically abused children rose from 311,500 in 1986 to 614,100 in 1996, a 97 % increase (Sedlack & Broadhurst, NIS-3, 1996). This increase in reports has not been accompanied by comparable increases in funding or staffing (Baumann, 1997).

Two alternative explanations for this increase in official reports of abuse must be considered. Either these numbers reflect a true increase in the incidence of child abuse and neglect within the population, or improved surveillance has resulted in better identification of maltreatment to the appropriate agencies. In the Executive Summary of the Third National Incidence Study, the authors propose that both of these dynamics

likely contributed to the increase in reports observed between the second national incidence study (NIS-2) in 1986, and the third (NIS-3) in 1996. The authors base this conclusion on the fact that dramatic increases occurred in both cases in which children were "endangered" and in those cases where children suffered serious injury (Sedlack & Broadhurst, 1996). The authors state, "It is unreasonable to suppose that quadruple the number of seriously injured victims of abuse and neglect existed at the time of NIS-2 and somehow escaped notice by community professionals" (p.14).

Whatever the cause of the increase in reports, CPS agencies must judge the validity of each allegation and respond accordingly. In 1997 twenty-five states reported data regarding reports that were "screened out", or not investigated. From this sample, 30.8% of cases were screened out, mostly due to characteristics of the case report that suggested that the situation was not within the purview of CPS responsibility (i.e. victim's age is over 18 years, a non-parent perpetrator is reported, report of juvenile delinquency, etc.) (DHHS, 1997). However, it seems that CPS is investigating fewer reports that do warrant agency attention, as well. For instance, Sedlack and Broadhurst note that CPS investigated only 28 % of reports of children meeting the "harm standard" in 1996, down from 44 % in 1986. The "harm standard" refers to children who have been determined to have suffered actual harm from abuse or neglect, as opposed to those who have been placed "at risk," but have not suffered observable harm (NIS-3, 1996).

The NIS data on these non-investigated cases are obtained by sampling other community professionals, such as sheriffs' departments, public schools, day care centers, mental health agencies, and others. Although the survey methodology used in the

National Incidence Studies could be challenged regarding the conclusions concerning population parameters, differences across NIS-2 and NIS-3 would appear to reflect real changes in reporting and investigative behaviors, because similar methods were used in each study.

A survey of CPS administrators and supervisors found that most of them report the existence of written agency policies regarding limiting investigations to those reports considered legitimate (meeting agency definitions of maltreatment). However, only 12% to 15% of supervisors reported that complaints that normally would have been investigated were screened out for investigation due to caseload burdens. (Downing, Wells, & Fluke, 1990). The authors of this study point out that supervisors acknowledging that legitimate reports were not investigated due to staff burdens could have negative implications, which may have resulted in underreporting. Despite this, at least one survey site in each state did report screening out reports due to worker overload (Downing, et al., 1990).

Child protection workers must make several other key decisions in addition to deciding which reports to investigate. The reports that are deemed appropriate for investigation are often prioritized in regard to the urgency of response warranted (Downing, et al., 1990). Upon investigation workers must make an immediate determination regarding the safety of the children in the home and whether removal of a child is required. These workers must decide which cases warrant ongoing social service intervention, and what kinds of services are likely to be most helpful. They must continue to assess the safety of the children while the CPS case remains open. Finally, they must

decide when it is safe and reasonable for CPS to terminate its contact with a family (Law et al., 1997; NCCD, 2000). Throughout the course of each case, the worker must assess, by some means, the degree of risk that a child will be re-abused, and the risk that he or she will suffer serious harm if protective measures are not taken.

Over the past fifteen years, CPS agencies have increasingly turned to formal "risk assessment systems" to assist caseworkers with the weighty decisions with which they are tasked. Here, "risk assessment systems" describes any formalized method of collecting or organizing information relevant to determining the likelihood of future abuse of a child. In a review of state CPS procedures, Pecora (1991) found that some agencies used the term "risk assessment" interchangeably with the term "needs assessment". This creates confusion regarding the purposes of these tools. A review of recent literature reveals that most references to "risk assessment" systems concern those that estimate, at least to some general degree, the likelihood of future abuse (Fluke, 1994; Johnson, 1993; NCCAN, 2000; Schene, 1996).

The focus of this review is only on those systems that attempt to quantify risk of subsequent physical abuse, either by estimating a specific probability of abuse for each case, or by assigning cases to ordinal categories of risk (i.e. high, medium, low). Systems that are designed primarily to assess risk of neglect, sexual abuse, or emotional abuse are not reviewed here because, although several risk variables have been associated with all three types of maltreatment (English, 1996), some factors, such as certain family characteristics, predict some abuse types but not others (Chaffin, 1996; Fuller, et al., 1998). For example, neglect is strongly associated with poverty and single parenthood,

while physical abuse occurs more often in families in which adult males are present

(Baird, 1999). Also, predictive factors may have different levels of importance (English,

1996) and interact differently in relation to abuse types (Kolko, 1998). For instance,

Kolko (1998) found that girls who were physically abused were older than were their

male peers, whereas, males who were sexually abused were older than were their female

peers.

Purpose of the Study

As noted above, CPS agencies have become deluged in an increasing flood of

reports of suspected abuse and neglect over the past two decades. Historically, workers

used the case study method to estimate the risk of future maltreatment. In this method,

the caseworker relies entirely upon his or her clinical experience and intuition to assess

future risk to the child (NCCD, 2000). Research over the past four decades has

consistently shown clinical judgement to be inferior to statistical methods in attempts to

make predictions regarding a variety of human behaviors (Dawes et al., 1989; Meehl,

1954).

During the 1980's CPS agencies were criticized for lacking a rational basis for

decision making (English, 1996). Although little research has been published regarding

the accuracy of clinical judgement in child welfare, results thus far have not been

impressive. Correlations between clinical predictions made by child welfare workers and

recurrence of maltreatment have not been found to be significant (Johnson, 1993). A

recent study compared case decision making among identified "CPS experts" and

caseworkers from four states. All participants read 70 case vignettes and made determinations regarding case disposition (e.g. foster care, in-home services, family preservation services). The researchers found wide disparity in decision-making, even among the "CPS experts" (Rossi, Schuerman, & Budde, 1996). The clinical expertise of these decision-makers proved to be inadequate for making consistent decisions in these complex matters.

A study of foster placement decisions made by child welfare workers based on clinical judgement revealed little consistency among workers regarding placement decisions, with workers likely to agree with one another only 25% above random chance (Lindsey, 1991). The author concluded that based on this low estimate of reliability, at least 48% of the placements were unnecessary and that 45% of the children needing placement were denied it (Lindsey, 1992). This estimate represents the best results that can be made when reliability is so drastically constrained (Ruscio, 1998).

In a study of caseworker decisions in Britain, Munro (1999) found that workers based assessment of risk on a narrow range of evidence, were biased toward information that was readily available to them, overly emphasized evidence that was vivid, concrete, or emotionally arousing (such as current, verbally-acquired information, as opposed to historical, written information). The workers were influenced by the order in which information was received, and failed to recognize the significance of known risk factors. Also, workers were found to be slow to revise their initial judgements in the face of mounting contradictory evidence. Munro concluded that errors in reasoning in child protection work are the products of human tendencies to simplify reasoning processes in

making complex decisions. The author recommends the development of means for checking intuitive judgement in a rigorous, systematic fashion.

Research done by the National Council on Crime and Delinquency in Rhode Island estimated that 15% to 25% of "high risk" cases are not opened for agency services, while many low risk families are carried on caseloads for months or even years (2000). The NCCD report states, "...the potentially grave consequences of error, the inherent difficulty in accurately assessing family situations and relationships, and the range of skills evident in CPS staff presents a near perfect equation for widespread disparity in case decision making" (2000, p. 2).

Although far from perfect, the formalized risk assessment systems being currently used by child protective agencies appear to offer the potential for improved prediction accuracy over the clinical judgements of individual caseworkers. Research on most of these models remains preliminary; however, reliability results on several of these models is encouraging. Few studies supporting predictive validity have been published; however, some models have demonstrated the ability to classify cases into different levels of risk with a fair degree of accuracy. English and Pecora (1994) sum up the primary argument for continued research into prediction of abuse recurrence when they state, " The reality is that CPS systems around the country are making critical decisions regarding the abuse/neglect of children absent empirical support or predictive models" (p.5). The limitations of current risk assessment instruments reviewed below should serve as a call for more and better research into prediction of abuse, not as an argument to abandon the effort.

Ruscio (1998) states that the largest impediment to child welfare decision-making is the poor reliability of clinical judgment. If a worker were to even devise a system of making his or her *own clinical judgments* more consistent (improved reliability) improvements in accuracy of predictions would be virtually guaranteed. Caseworkers need tools to help them to make more accurate assessments of risk.

Although risk assessment systems currently used in CPS agencies are an improvement to unaided caseworker judgement, their predictive ability has proved to be marginal. Several obstacles have impeded efforts to predict recurrence and severity of child abuse. These obstacles include selecting the appropriate criterion variable, difficulties related to predicting rare events (low base rate), statistical analysis problems caused by known and unknown interactions among predictor variables, and concerns of racial/cultural bias in instrument development and application.

In light of the limited success of current risk assessment models to predict recurrence and severity of child abuse, the proposed study explores the potential of computer-based learning systems, known as artificial neural networks (ANNs), to improve predictive accuracy in child abuse risk assessment. ANNs have demonstrated utility in prediction and classification problems, especially in situations in which underlying relationships among predictor variables are unknown (Garson, 1998), when data are corrupted by random error or missing data (Hartzberg, Stanley, & Lawrence, 1990; Lippman, 1987; Weiss & Kurlikowski, 1991), when relationships between predictors and outcomes are nonlinear (Gallinari, Tthiria, & Goglman-Soulie, 1991; Gordon, 1992), when assumptions of multivariate normality do not hold (Garson, 1998),

and when base rate of the outcome of interest is low (Gordon, 1991). These conditions often apply to CPS databases that are used to develop and test risk assessment models.

The proposed study consists of a retrospective exploratory analysis of a large sample of substantiated cases of child physical abuse reported to the U.S. Air Force Central Registry between January, 1990 and June, 2000. Generalizing beyond the Air Force population to other populations is problematic due to differences between the populations (e.g. age, employment status, mobility). However, given the unique characteristics of the military, it is important to investigate child maltreatment recurrence and severity within this environment and to lay a foundation for the development of risk assessment tools to help those charged with preventing and responding to such incidents within Air Force communities.

Objectives

Specifically, the objectives of the proposed study are:

1. To systematically develop and train a neural network model from the case data to predict recurrences of physical child abuse.

2. To test the recurrence prediction model on a cross-validation data sample to counter artificial inflation in accuracy rates due to potential overtraining of the network.

3. To compare the predictive accuracy of the neural network model to results produced by an alternative method (binary logistic regression).

CHAPTER II

REVIEW OF THE LITERATURE

Empirically-Supported Predictors of Recurrence of Physical Child Abuse

A considerable amount of research has been conducted with regard to identifying correlates of child abuse; that is personal, situational, and environmental conditions that coincide with the occurrence of abuse. Gil (1970) is generally credited with conducting the first study of child abuse in the U.S. Examining case characteristics from a sample of child abuse reports received in 1967-1968, Gil concluded that abuse was a multidimensional phenomenon produced by environmental chance factors, environmental stress factors, deviance of caretakers and children, and disturbed intrafamilial relationships. While subsequent authors have criticized many of this study's conclusions (see Daley & Pilavin, 1982; Seaburg, 1977), this multidimensional conceptualization of the causes of abuse has continued to guide research in this arena.

CPS agencies are not primarily concerned with identifying families that are likely to *initially* abuse their children, but whether those families already suspected of, or confirmed for abuse, are likely to *re-abuse*. Factors that are predictive of initial abuse are not necessarily those factors associated with re-abuse. In fact, because those families that are being assessed by CPS are already suspected of committing abuse, they are likely to have many factors associated with initial abuse. The task is to predict which subset of this population is likely to re-abuse (Wald & Woolverton, 1990).

This review examines factors that have received empirical support in relation to the outcomes of recurrence of child physical abuse. Factors found to be associated with

initial abuse, but not with re-abuse, are excluded. Although studies aimed at identifying

correlates of only initial incidents of physical child abuse are not included here, it should

be noted that research into correlates of abuse recidivism has drawn upon studies of

initial abuse to guide variable selection (Johnson & L'Esperance, 1984). Included

variables are typically categorized as parent/perpetrator characteristics, child/victim

characteristics, and environmental/situational characteristics. Reviewed below are those

factors from these three domains that have been found to correlate with the outcome of

repeated abuse (as defined as repeated referrals or repeated substantiated reports). Factors

that have shown mixed results in regard to this correlation will also be included, as they

may be of interest for inclusion in future predictive models.

Predictors of Re-abuse

Child characteristics. Child characteristics that have received research attention

are age, gender, race, and special needs (i.e. handicapping condition). The child's age at

the time of abuse has been considered in nearly all studies of abuse recurrence. Applying

survival analysis to 24,507 child maltreatment cases received in Colorado from 1986-

1989, Fryer and Miyoshi (1994) found the child victim's age to correlate with recurrence

of all maltreatment categories, with younger children more likely to be repeatedly abused,

and substantially fewer repeat cases involving adolescent children. A subsequent study

applying similar methods, but examining national data from the National Child Abuse

and Neglect Data System (NCANDS), failed to find increased risk for the youngest age

group, but did find significantly lower recidivism in regard to the oldest age group (12-17

years) (Fluke, Yuan, & Edwards, 1999). Weedon, Torti, and Zunder (1988), not

differentiating between abuse and neglect, examined 147 families using the Family Risk Assessment Matrix. Fourteen variables, including child's age, were found to be predictive of a subsequent founded report. The 1985 National Violence Survey, a telephone survey of 6,002 households that questioned respondents about abusive behaviors, found a curvilinear relationship between child's age and risk of suffering both minor and severe violence, with preschool aged children being at higher risk than either toddlers or adolescents (Wolfner & Gelles, 1993). There is ample evidence for the child's age as a predictor of re-abuse to justify consideration in future model-building efforts. This factor has also been found to correlate with initial abuse (Wolfner & Gelles, 1993).

Child's gender has been found to correlate with initial occurrences of abuse, girls being at higher risk of experiencing abuse (Sedlack, 1997) and with injury severity, boys being at greater risk for severe injury (Rosenthal, 1988; Sedlack & Broadhurst, 1996). The role of gender in *recurrence* of physical abuse is not clear. In the Colorado study, female children were found to be somewhat more likely to be revictimized. Age also interacted with gender, with girls from ages 1-6 years being at the greatest risk (Fryer & Miyoshi, 1994). In other studies child gender has not been reported to correlate with re-abuse.

Several studies have examined the child's or perpetrator's race in relation to re-abuse. African American families appear to be under increased scrutiny, in that a disproportionate number of these families come to the attention of CPS, despite national incidence studies that estimate essentially equal prevalence of maltreatment in Caucasian and African American populations. However, the effect of race on re-referral, as opposed

to initial reporting, is a separate matter for consideration. Fluke, Yuan, and Edwards (1999) found mixed patterns across states in regard to time to abuse recurrence and race, with Caucasian families recidivating sooner in some states, and African American families in others. Asian American families consistently demonstrated a longer time until recurrence. The Children's Research Center of the NCCD conducted a study that examined relationships between race and assigned risk ratings, and between race and re-abuse rates in Georgia, California, and Michigan, three states that have adopted actuarial risk assessment models. This study found no significant differences across racial groups (white, African American) (California also included Hispanic) in regard to the initial risk level assigned using the instrument, and no significant differences in subsequent recurrence of maltreatment (Baird, Ereth, & Wagner, 1999).

The authors of the CRC study point out that the use of separate abuse and neglect risk assessment tools likely addressed any potential problem with racial bias in the study. Because, while overall maltreatment rates are similar for whites and African Americans, there is great variation across abuse types. African American families have a higher rate of neglect reported; White families have a higher rate of abuse reported. Because neglect is strongly associated with poverty and single-parent homes, and abuse (physical and sexual) is most often committed by males, race appears to be a flag of other influential socioeconomic factors in this context. It is unclear how much interactive effect exists between race and other demographic variables (English & Pecora, 1994). It would seem important to include race in future predictive models to allow for such interactions and to examine prediction accuracy across racial groups.

14

A few studies have examined the presence of special medical or behavioral characteristics of the child as a potential predictor of abuse. Weedon, Torti, and Zunder (1988) found an increased risk for maltreatment across abuse types for children with impaired physical or mental abilities. Burrell, Thompson, and Sexton (1994) surveyed mothers of children with (n=53) and without (n=60) special medical, psychological, or emotional problems. Mothers of "special needs" children reported significantly higher levels of child abuse potential. In a recent analysis of 120,000 referrals received in a Northwest CPS system, researchers found that children with developmental disabilities were overly represented in "multiple-referring" families (Marshall & English, 2000). It appears that "special needs" children may experience increased risk for re-abuse. However, few studies to date have examined this variable. The presence of special medical, psychological, or cognitive conditions deserves further examination.

Caregiver/perpetrator characteristics. Characteristics of the perpetrators, parents, and/or caregivers that have received research attention are gender, age, race, domestic violence, childhood history of abuse, and substance abuse. Additionally, several measures of psychological well being, parenting skills, and coping skills have been examined in relation to subsequent abusive behavior.

National incidence studies have shown that male caregivers perpetrate most cases of physical abuse, 67% of cases identified in 1995 (Sedlack & Broadhurst, 1996). However, the psychological characteristics of mothers in related to abuse has received have received more research attention than have those of fathers. This may be partially explained by the fact that, when neglect is considered in combination with abuse, mothers

comprise the majority of identified perpetrators. Rather than examining the relative risk of re-abuse as a function of the gender of the perpetrator of the initial incident, research has focused on the differential effects of other risk factors, such as substance abuse, for mothers and for fathers, or have looked at factors that increase risk of subsequent abuse specifically for mothers.

Mother's age has been examined as a correlate of physical child abuse. Children of younger mothers (operationalized as either <18 or <20 years old across studies) have been found to be at increased risk for initial incidence of physical abuse (Creighton, 1985; Zuravin, 1987) and for recidivism of physical abuse (Herrenkohl & Herrenkohl, 1979). Zuravin (1988) re-examined this relationship while statistically controlling for chronic sociodemographic stress, as measured by number of live births, life history of unemployment, and low educational achievement. This study revealed that the number of live births was a particularly strong mediator variable. Mothers who birthed several children beginning at a young age were particularly vulnerable. An analysis of the 1985 national violence survey revealed that mother's age at the time of the child's birth was predictive of future physical abuse, however, mother's age at the time of the abuse incident was not predictive (Connelly & Straus, 1992).

As discussed in regard to child characteristics, race has been examined as a predictor of recurrence of abuse. Some studies have examined race of parents, versus race of the victim. And, however operationalized, as noted above, while there are no differences in overall rates of occurrence and recurrence of maltreatment (Sedlack & Broadhurst, 1996; Baird, Ereth, & Wagner, 1999), significant differences do exist

between White and African American families in relation to abuse type (Baird, Ereth, & Wagner, 1999).

It has been reasonably assumed that many child abusers were themselves abused as children and that abusive behavior is transmitted intergenerationally. Gil (1970) reported that 14% of mothers and 7% of fathers in his sample of abusers reported a childhood history of abuse. Shapiro (1979) reported that 16 percent of known abusers were abused as children. A perpetrator's childhood history of abuse has been found predictive of recidivism of abuse in recent studies (English, 1996; English & Pecora, 1994). Researchers in Washington applied multiple variable selection techniques in order to construct a parsimonious model to predict single and multiple recurrences of re-abuse. Perpetrator's childhood history of abuse was identified as one of eight key variables in the final model. Through observation of neural network weights, the authors also noted that this factor appeared to interact in complex ways with other factors such as victimization of others and chronicity of child abuse and neglect (Marshall & English, 2000).

A history of domestic violence (spouse abuse) has been found to be strongly associated with the occurrence of child physical abuse. Ross (1996) found that the greater the amount of violence against a spouse in a given family, the greater the probability of a child suffering physical abuse. This relationship was stronger for husbands than for wives. Domestic violence was identified as one of five key predictive factors of recurrent child abuse by researchers in Washington state (English, Marshall, Brummel, & Orme,

2000). DePanfilis and Zuravin (1999) found that the presence of partner abuse reduced the time until a subsequent child abuse incident during the course of treatment.

Parents with substance abuse problems are more likely to initially abuse and neglect their children (Chaffin, Kelleher, & Hollenberg, 1996). Wolfner and Gelles (1993) found that parents who self-reported even one instance of using illegal drugs were more likely to abuse their children. Both mothers and fathers with histories of substance abuse disorders report elevated levels of abuse potential (Ammerman, Kolko, Kirisci, Blackson, & Dawes, 1999). Additionally, substance abuse has been found to be predictive of re-referral and substantiated recurrence of child abuse (English, Marshall, Coghlan, Brummel, & Orme, 2000).

Several psychological measures have been used for the purpose of predicting recurrence of abuse. The Child Abuse Potential Inventory (CAP or CAPI) (Milner, 1979) is the most widely used and empirically supported self-report measure of physical child abuse potential. Scores on this measure are occasionally used as a dependent measure, as opposed to measuring actual incidents of abuse (e.g. Ammerman, et al, 1999). Other measures of such constructs as parenting skill, stress, depression, parents' expectations of their children have been used as independent predictors of abuse.

Johnson and L'Esperance (1984) found that the mother's parenting skill and reasonableness of her expectations of her children, as measured by caseworker ratings on Likert-type scales, were predictive of subsequent physical abuse. DePanfilis and Zuravin (1999) found that a "maternal problems" construct, measured by summing three-item indices rated by case reviewers of alcohol problem, drug problem, and problem solving

deficits was predictive of recurrence of abuse. Chaffin, Kelleher, and Hollenberg (1996) found that mothers with a diagnosis of depression were 3.45 times more likely to physically abuse their children during a one-year follow-up period than were their non-depressed counterparts.

Situational/Environmental Factors. Socioeconomic status has consistently been found to correlate with both physical abuse and neglect, with children from families with an annual income of less than $15,000 being at 22-25 times more likely to suffer some type of maltreatment (Sedlack & Broadhurst, 1996). Wolfner and Gelles (1993) found that children of parents who were unemployed or who worked "blue-collar" jobs were at increased risk of abuse. In regard to re-abuse, Levy and colleagues found that residence in public housing, having unmarried and/or unemployed parents, and being Medicaid recipients placed children at increased risk for revictimization (Levy, Markovic, Chaudhry, Ahart, & Torres, 1995). Inkelas and Halfon (1997) also found a positive relationship between receipt of income support and subsequent abuse.

Strong evidence exists to support the conclusion that children in larger families are more likely to experience physical abuse. Family size has been operationalized in various ways across studies. Johnson and L'Esperance (1984) found that the presence of more than one child in the home increased risk of recurrence of physical abuse. DePanfilis and Zuravin (1999) used a dichotomous measure, with three or more children, in combination with other factors, constituting increased family stress. Zuravin (1988) examined the combined relationship of teenage first births and subsequent number of live births. Regardless of these operational variations, the number of children in the household

has been consistently found to positively correlate with physical abuse (Baird, 1988;

Connelly & Straus, 1992; DePanfilis & Zuravin, 1999; Johnson & L'Esperance, 1984;

Zuravin, 1988). A notable exception is Wolfner and Gelles's analysis of the National

Family Violence Survey, in which they found a curvilinear relationship between number

of children and severe and minor violence, with a peak at four and five children,

respectively (1993).

Conclusions

　　　Empirically supported correlates of abuse recurrence are typically categorized as

child/victim characteristics, parent/perpetrator characteristics, and

environmental/situational characteristics. Child characteristics that have received at least

some degree of empirical support as correlates of recurrence of physical abuse are age,

gender, race, and handicapping conditions. Perpetrator/Parent characteristics that have

been found to correlate with recurrence of abuse are gender, age, domestic violence, a

childhood history of abuse, and substance abuse, as well as parenting deficits, poor

coping skills, and experience of stress and depression. Supported situational correlates of

re-abuse are poverty and family size.

Types of Risk Assessment Systems

　　　Most states have implemented one of two alternative types of risk assessment

systems designed to minimize the risk of future harm (Berkowitz, 1991). Two broad

categories of the risk assessment models commonly used are "consensus-based systems"

and "actuarial systems". Consensus-based systems are developed by the consensus

judgement of experts, and are comprised of items generally thought to be associated with risk of future maltreatment. In these systems, workers assess specific client characteristics identified by experts, then exercise their own clinical judgement about the risk of future abuse (NCCD, 2000). The Washington Risk Assessment Matrix (WRAM or WARM) is one of the more widely used and researched consensus-based models. It groups risk factors by child characteristic (5 items), severity of abuse/neglect (9 items), chronicity or recurrence of episodes of child abuse/neglect (1 item), caregiver characteristics for primary and secondary caregivers (11 items each), caregiver-child relationship (6 items), social economic factors (4 items), and perpetrator access. The risk factors are rated 0-5, for no risk, through high risk. A summary assessment using these characteristics is obtained (Baird, et al., 1999).

Actuarial systems are developed through empirical study of CPS cases to identify factors with a strong association with abuse. Workers score these items to identify a level of risk for each (NCCD, 2000). The Michigan Family Risk Assessment of Abuse (FRAAN) represents a typical actuarial risk assessment system. It contains separate neglect (11 factors) and abuse (12 factors) assessments, along with a separate caretaker strengths and needs assessment (13 factors). Caseworkers score each individual item on the neglect and abuse scales, total the results, and rate the family into one of four categories of risk (Baird, et al., 1999).

Considerable progress toward the development and validation of such tools has been made over the past twenty years: Reliability and validity studies of consensus-based and actuarial systems are reviewed below. However, several obstacles have impeded

efforts to make accurate predictions about which children are most likely to suffer further abuse.

Research Problems in the Prediction of Physical Abuse

Numerous problems have hindered researchers' attempts to predict which children will be abused. Among these are what Parton (1997) refers to as the "definitional fallacy" and the "statistical fallacy". The "definitional fallacy" concerns ambiguity and disagreement about what constitutes child abuse. In prediction efforts this translates into discrepancies in definitions of the dependent (or outcome) variable of interest. While this article seeks to simplify this matter by limiting its scope to physical child abuse, it will be seen that in studies of risk assessment instrument performance, researchers have defined outcome variables is various ways. The "statistical fallacy" to which Parton refers, is the inevitability of making "false positive" and/or "false negative" predictions when attempting to predict any phenomenon that has a low rate of occurrence (base rate) within the population of study. When attempting to predict a phenomena that occurs at a low base rate, the researcher may choose to minimize false positives, at the cost of increasing false negatives, or vice-versa (Murphy-Berman, 1994). This decision becomes a value judgement, driven by the researcher's priorities and intended use of the information obtained. Problems regarding low base rates will be discussed in more detail below.

Another obstacle to accurate prediction of child physical abuse lies in defining the necessary independent (predictor) variables to be included in the instrument or model. Research has made progress in identifying a number of variables that are related to the

recurrence and severity of physical child abuse. However, what are less clear are the interrelationships among these predictor variables. Several authors have commented on problems associated with anticipated interactions among variables associated with abuse (English 1994; Sedlack, 1997; Wald & Woolverton, 1990). The nature of these complex relationships is not well known at present. Scoring strategies and statistical methods used in most prediction models are not optimal when such complex relationships exist among variables, when base rates are low, and assumptions of multivariate normality and equal variance are not met (Lyons, Doueck, & Wodarski, 1996; March, 1998).

Defining the Criterion Variable

Comparisons across studies of risk assessment are difficult due to differences in the dependent (criterion/outcome) variable used, that is, assessing the risk of "what". First, it must be determined if the intent is to assess the risk of "first abuse" or "re-abuse". The correlates of first abuse may differ substantially from those of re-abuse (Lyons, et al., 1996). Prediction of initial abuse is particularly difficult due to uncertainties about the true rate of abuse in the general population and difficulties in establishing relevant control groups (i.e. known non-abusers). Although the National Incidence Studies mentioned above represent the most comprehensive attempt to estimate abuse and neglect within the United States, they still cannot provide reliable estimates of abuse in the general population. Although the NIS samples various community sources in addition to CPS agencies, they fail to capture instances of maltreatment that go unobserved by *any* community professionals. However, CPS agencies are tasked with responding to situations where maltreatment has already "allegedly" occurred. Therefore, risk

assessment in the context of CPS work is generally concerned with the risk of recurrence of maltreatment (re-abuse) (Johnson, 1993; Wald & Woolverton, 1990). While of significant social importance, prediction of initial abuse is not of primary practical concern to CPS agencies, and is beyond the scope of this paper.

Studies of risk assessment related to physical child abuse have defined outcome variables differentially as: subsequent re-referrals of abuse (Johnson & L'Esperance, 1984; Marks & McDonald, 1989; Weedon, Tori & Zunder, 1988), subsequent substantiated cases (Baird, 1988; Johnson & Clancy, 1988; Wells & Anderson, 1992; Zuravin, Orme, & Hegar, 1991), number of child injuries and subsequent foster care placements (Baird, et al., 1995), clinical judgements or activities of CPS workers (Doueck, English, & Moote, 1993; Fanshel, et al., 1994; Johnson & Clancy 1988; McDonald & Marks, 1991), and injury severity (Paddock, 1995; Zuravin, Orme & Hegar, 1994).

Research regarding how individual caseworkers and CPS agencies use prediction or classifications produced by risk assessment systems in decision-making processes is important. After all, if this information is disregarded, there is little point in requiring workers to complete the additional paperwork. However, the practical utility of these tools largely depends on their ability to produce valid assessments of risk. Until the validity of these instruments is established, research regarding their organizational applications seems premature. At this point, empirical evidence supporting predictive validity of the models currently in use is mixed. Therefore, continued research aimed at testing the predictive validity of risk assessment systems deserves a high priority.

Toward this goal, relevant outcome variables are those that indicate "recidivism" or recurrences of abuse in cases that have come to the attention of CPS. Whether future "referral" or future "substantiated referral" provide the best test of validity is debatable. Is a future referral that is not substantiated an instance of "recidivism"? Perhaps those who have had initial CPS contact are under greater surveillance, resulting in a lower threshold for further inquiries regarding suspicion of abuse.

Limiting the outcome variable to only those re-abuse incidents that are subsequently "substantiated" may fail to adequately capture the outcome of interest (actual recurrence of abuse). Institutional variables have been found to influence substantiation decisions. For instance, some states distinguish only between two categories, "founded" versus "unfounded" (two-tier states), whereas other states include a middle category, such as "indicated" or "inconclusive" (three-tier states) (English, et al., 2000). Substantiation rates, which are about 40 % nationally, are typically higher in two-tier versus three-tier states (Drake, 1996; Flango, 1991; Wells & Anderson, 1992; Zuravin, et al., 1995). Drake (1996) conducted a review of a variety of studies on substantiation. Evidence suggests that rates of substantiation vary by site, policy, law, and level of risk.

Drake (1996) challenges the use of substantiation decisions as outcome variables in research related to CPS practice, especially the use of a finding of "unsubstantiated" as a proxy for frivolous reporting or unnecessary CPS intervention. The author points to research that shows that many cases are unsubstantiated due to lack of evidence, despite the worker's belief that maltreatment has occurred (DHHS, 1981 [NIS-1], 1996;

Giovannoni, 1991; Jason, Andereck, Marks, & Tyler, 1982; Zuravin, et al., 1995). The substantiation finding is "evidence dependent" and may be unrelated to the severity of harm suffered or to the need for social service intervention. The author concludes that current empirical data suggest that the majority of unsubstantiated reports include either service need, harm to the child, or maltreatment that is not demonstrable (Drake, 1996, p.267).

English and colleagues (2000) at the State of Washington Office of Children's Administration Research recently conducted a large, state-wide, cross-sectional study of causes and consequences of CPS workers' substantiation decisions. The authors used an artificial neural network and statistical analyses to examine determinants of substantiation decisions that were made on 12, 871 referrals. Factors most influential of caseworker substantiation decisions were: chronicity of abuse, factors related to incident severity, caretaker's recognition of the problem, parenting skills and substance abuse, child's behavior, and fear of the caretaker. Cases that were referred by law enforcement were found to have a higher probability of substantiation, over and above the cumulative effect of the risk factors. Interestingly, the rate of re-referral for founded and unfounded cases were not significantly different. This suggests that substantiation status may not be a useful criterion for service planning decisions if the goal is to reduce the incidence of re-abuse.

It appears that determinations regarding case substantiation are influenced by observable physical evidence (associated with "severity" of abuse) by statements made by the alleged victim (influenced by the age of the child), by the confessions of the

perpetrators of abuse and by differing standards across state agencies (English, et al., 2000; English & Pecora, 1994). It has also been argued that substantiation decisions are biased against racial minority groups. African Americans are more likely to have maltreatment allegations substantiated in spite of the fact that national incidence estimates indicate that there are no differences in maltreatment rates for African Americans and Whites (Baird, 1999).

The use of re-referral as a criterion has received similar criticism. Cases in which re-referral is likely are not necessarily those in which a child is at risk of suffering serious harm (English & Pecora, 1994). Severity of injury resulting from abuse as an outcome variable has received less empirical study (Daley & Pilavin, 1982, Hampton, 1987; Paddock, 1995; Rosenthal, 1988; Seaberg, 1977; Zuravin, et al, 1994). The use of injury severity as a criterion variable adds value to the research base because CPS staff must be concerned with not only *if* further abuse will occur, but also, which children are at risk of suffering *serious harm* (Wald & Woolverton, 1990). Essentially, by using injury severity as the outcome variable, these studies are not capturing recidivism, per se, but rather a separate, but related construct. The literature generally refers to this as "safety assessment". Safety assessments go beyond predicting the potential of maltreatment at some point in the future to suggesting that future maltreatment may be severe, and thus may require immediate intervention (DePanfilis & Scannapieco, 1994). While risk of suffering any further abuse and the risk of suffering serious physical injury are probably closely related, they are not identical. Factors that are predictive of one are not necessarily predictive of the other, but these factors may overlap to a significant degree.

A prediction that a child is at high risk for continued abuse at a mild severity level is quite different from a prediction that the child is at high risk for suffering substantial physical harm. These different types of predictions would suggest different interventions by CPS. The latter is more likely to warrant removal of the child from the home. Consequently, many states now differentiate "risk assessment" from "safety assessment", with safety assessment referring to the likelihood of imminent and serious harm (DePanfilis & Scannapieco, 1994), and risk assessment referring to the likelihood of recidivism. Separate instruments have been developed for each of these assessments in some state agencies (Meyer, 1998), whereas some risk assessment models incorporate a safety assessment component (Holder & Corey, 1993).

The use of either rates of all re-reports, or for substantiated reports only can be reasonably argued as the best measure of the predictive power of risk assessment systems. Since CPS must assess the validity of each report received, any additional report of abuse taxes agency resources, and may indicate ongoing abuse even if evidence is lacking to confirm the allegation. Limiting the criterion to only "substantiated" re-referrals reduces the effects of increased surveillance for previously identified families and the effects of spurious reports, but also misses subsequent abuse that did occur but cannot be confirmed through evidence or admission. It seems unnecessary to chose between these outcome measures when information regarding both is available from CPS files. Comparability of results from future studies of the risk assessment systems performance will be enhanced if researchers agree to report on *both* re-referral and re-abuse (substantiated). Additionally, research regarding the prediction of injury severity

(safety assessment) has been minimal. Much more research in needed to improve prediction of serious harm, as opposed to prediction of recurrence of abuse across all levels of severity.

Sensitivity versus Specificity

Prediction systems must address both sensitivity (correctly identifying families where maltreatment will recur) and specificity (correctly identifying families in which maltreatment will not occur). Low sensitivity results in "false negative" predictions, while low specificity results in "false positive" predictions.

Prediction of rare phenomena is difficult. This is typically referred to as a "low base rate" problem. This is an issue not only in child maltreatment studies, but also in the prediction of violence, in general. Optimal increases in prediction occur when the base rates are 50%, which means that 50% of the sample are criterion cases (abusers). As the base rate decreases, the percentages of false positives and false negatives will change dramatically (Milner & Campbell, 1995). For example, assuming a hypothetical base rate for child abuse in the general population of 5%, in a group of 100 subjects, five would be abusers. A classification system that is 80% accurate would correctly classify four of the five abusers. However, it would also label 19 non-abusers as abusers (Milner & Campbell, 1995). In fact, given a base rate of 5%, one could achieve 95% accuracy by simply classifying all of the subjects as non-abusers.

As discussed above, risk assessment systems used in child protective services are primarily targeted toward prediction of re-abuse in those families that have already been identified to the agency. The base rate for re-abuse in these families is considerably

higher than the base rate of initial abuse in the general population. Estimated maltreatment recurrence rates have ranged from 18% to 60% (Doueck, English, DePanfilis & Moote, 1993; Wald & Woolverton, 1990). A recent analysis of the National Child Abuse and Neglect Data System (NCANDS) revealed an average rate across states of "substantiated" re-abuse cases at 19% at 12 months observation. (Fluke, et al., 1999).

Researchers can choose to maximize the sensitivity of a test instrument, at the cost of reducing specificity, meaning that more correct predictions of re-abuse would be made, but at a cost of incorrectly labeling non-recidivating families as recidivists. Errors in the opposite direction could also be minimized, resulting in a lower rate of identification of recidivist families. Situations where one is attempting to predict such rare phenomena with less than perfect measures require value judgements. These decisions must be driven by research and practice priorities. Is it worse to incorrectly label non-recidivists as recidivists, or to miss cases in which abuse will recur? Obviously, incorrectly labeling even a relatively small number of people as abusers could create serious civil and emotional consequences and failure to predict recurrences of abuse will place children in jeopardy.

In medicine, social work, and other helping professions higher levels of false positives are typically tolerated, because false positive cases can be re-evaluated at some later time, but false negatives may never be seen again until some serious morbidity occurs (Comasso & Janannathan, 1995). For a predictive instrument to be used as a formal screening device in this arena it must have high specificity and extremely high sensitivity (Wald & Woolverton, 1990).

Baird, Ereth, and Wagner (1999) argue that overall prediction accuracy is not necessarily the best standard by which to gauge risk assessment instruments when probability levels are different from 50-50. As discussed above, when attempting to predict rare events, simply assuming that an event will not occur may result in better accuracy than any attempt to determine where or when occurrence is likely. If subsequent maltreatment is reported in only 15% of open CPS cases, then simply predicting no re-referral will result in an 85% "hit rate". But, while highly accurate, such a prediction is of little value to CPS. The "sensitivity" of the prediction is .85, but the "specificity" in this case; the identification of those who re-abuse is 0.0. It is possible that a valid risk assessment system could result in a higher percentage of false positives and false negatives and still provide the agency with higher quality information about the relative probability of subsequent maltreatment (p. 9).

Baird demonstrates that an actuarial model that achieves an 82 % "hit rate" (3% lower than when no re-occurrence is predicted) would provide more valuable data, in that it would correctly identify 11 of 15 cases in which maltreatment did recur, although while being incorrect in 56% of cases where recurrence of abuse was predicted (false positives) (p. 9, 1999). Clearly, with such a model one could not assume that those identified as "high risk" would re-abuse (56% would not). But, it would "miss" only 27 % of cases in which abuse did recur. The ratio of re-abuse in the high-risk group to the low risk group is more than 8:1. This is information that could be used to inform *some* CPS decisions.

Other fields in which low base rate phenomena are of interest, such as medicine, juvenile justice, and adult corrections have abandoned the idea that risk assessment is an

exercise in prediction (Baird, 1999). Instead, "high-risk" simply denotes inclusion in a group with significantly higher rates of recidivism. As will be seen upon review of the prediction accuracy achieved thus far in the area of physical abuse, precise predictions regarding re-abuse for a particular case in not yet feasible.

Interactions among Predictor Variables

Interactions Identified thus far. In statistics, "interaction effects" refers to situations in which changes in dependent (outcome) variable are different than would be predicted based on the main effects of the independent (predictor) variables, due to influential relationships among the independent variables (Norusis, 1997). Numerous researchers have remarked on the presence of interaction effects among correlates of abuse recidivism (Ammernan, 1999; Daley & Pilavin, 1982; Howing, Wodarski, Kurtz, & Gaudin, 1989; Hutchinson, 1989; Milner, 1995; Parton, 1997; Pecora, 1991; Sedlack, 1997; Schene, 1996; Zuravin, et al., 1994;). Several such interactions have been identified. Kolko (1998) found the variables of gender and age of the child to interact in relation to the differential probability of being physically or sexually abused. Girls who were physically abused were older than their male counterparts, whereas males who were sexually abused were older than were their female peers. Kruttschnitt et al. (1994) found an interaction between race and poverty in relation to severity of abuse, with current poverty becoming non-significant when the effect of race is statistically controlled. In a test of a model predictive of child injury using case report characteristics, Zuravin et al. (1994) found interactions between child's age and mother's age, and between child's age and child's gender. Older children were likely to suffer less severe injury, and the older

the mother, the stronger this inverse relationship. Age did not affect the severity of injury

to girls, however injury severity was greater for younger boys than for older boys. Gelles

(1989) discovered an interaction between caretaker gender and poverty as predictors of

future child abuse, with violence being a function of poverty in mother-only homes, but

unrelated to income among single fathers. Ross (1996) found that men who were abusive

toward their spouses were more likely to commit child abuse than women who were

abusive toward their spouses (gender x domestic violence interaction). DePanfilis and

Zuravin (1999) found that levels of family stress interacted with social support deficits in

relation to child maltreatment recurrences, with adequate levels of social support acting

to reduce the time until recurrence of abuse in families experiencing high levels of stress.

Ammerman (1999) found that substance-abusing fathers were more likely to self-report

abusive potential than were substance-abusing mothers (gender x diagnosis interaction).

Research Problems Related to Interactions. Although knowledge of interactions

among predictors of child abuse is growing, many more unidentified interactions likely

exist (Pecora, 1991). Some factors may be more important than others, some may be

multiplicative, some additive. At present these relationships are unknown and it is

extremely difficult to conduct research that has a sample size large enough for statistical

analysis of such interactions (Wald & Woolverton, 1990).

Instrument scoring procedures and data analytic techniques that fail to account for

the interactions among selected predictors will produce flawed predictions or

classifications. Adding intercorrelated risk factors may result in an overestimate of risk

(Wald & Woolverton, 1990). Most risk factor scoring systems are unable to adjust for

how factors interact and non-linear relationships, in part because of a lack of empirical

data regarding which interactions are most important (Pecora, 1991). Multicollinearity of

predictor items decreases the chances of any one predictor obtaining statistical

significance in multivariate analysis, which can lead to the erroneous conclusion that

none of the predictors are important (Fuller & Wells, 1998).

Use of Risk Assessment at Various Decision Points

CPS workers make numerous decisions that affect the safety and welfare of

children. One can visualize a continuum of contact starting with the initial referral,

requiring a decision regarding whether or not the report warrants investigation. If

assigned for investigation (or assessment) a decision regarding "substantiation" must be

made. That is, does available information support the conclusion that a child has been

abused, as defined by the state's applicable statutes? Some states use alternate

terminology, such as "founded" rather than "substantiated", but the idea is essentially the

same, whether or not abuse can be confirmed through evidence or admission. CPS staff

must decide which intervention program(s), if any, to offer to the family. When services

are provided, the case must be monitored for progress. Finally, a determination must be

made regarding when CPS may discontinue involvement with the family without placing

a child at risk. Also, at any point, but usually during the initial assessment, the

caseworker must decide if removal of a child from the home is necessary due to risk of

imminent harm.

To date, risk assessment research and model development has focused primarily

on deciding whether to continue to provide service after conducting an initial

investigation of maltreatment (Johnson, 1996) and, to a lesser degree, on "screening" of

referrals from initial report characteristics to determine if, and when, the report should be

investigated. One reason for the emphasis on the "front end" of the process is that this is

where the greatest number of caseworker decisions occurs (Law, Maumann, Gober,

Schultz, Ohmart, & Kern, 1997). A lesser amount of research has been conducted into

the other decision points. "Safety assessment", determining the risk of "imminent and

serious harm" has received somewhat less research attention (DePanfilis & Scannapieco,

1994; Fluke, Edwards & Wells, 1996; Salovitz, 1992;).

Risk assessment beyond the initial assessment phase of the case management

continuum has received very little study (Baumann et al., 1997). This may be partly due

to greater agency interest in those factors and decisions that present during initial

assessment of the case. But lack of research attention to these later phases may also be

due to the increased complexities that arise once treatment has begun. Numerous

additional factors related to intervention become relevant and these new factors are likely

to interact in complex ways with those factors that were assessed during the initial

investigation. The question changes from "which families are most likely to re-abuse?" to

"given "x" intervention, which families are most likely to abuse?" Wald and Woolverton

(1990) propose that *this* is the real question of interest in child protective work, and state

that no studies have examined the impact of services on reducing risk.

Some research has attempted to statistically control for treatment effects (typically

defined by duration and intensity of caseworker contact). The results of these studies

have been mixed. Johnson and L'Esperance (1984) found a significant relationship

between recurrence of abuse and duration of services, with recurrent abuse less frequent in cases that remained open longer. A subsequent study, also done by one of these investigators, failed to support this relationship (Johnson & Clancy, 1989). However, the *intensity* of service contact was found to affect the recurrence of maltreatment in other studies (Johnson, 1995; Lutrell, Hull, & Wagner, 1995). Fluke and colleagues (1999) found that cases that received postinvestigation services were more likely to experience re-abuse compared with those who did not. But these studies have focused on quantity and duration of services, without specific attention to the content of these services. The differential effects of various intervention programs, and combinations of these programs, in relation to those case report characteristics thought to be important, have not been explored. Such research would have to address not only those obstacles discussed above, but also many others. For instance, ethical questions regarding the viability of randomly assigning subjects to treatment and control groups, problems of attrition of participants, and fidelity of treatment provided would have to be addressed. Ultimately, this type of research will be invaluable to CPS agencies, but it is not yet feasible due to the limited state of knowledge in regard to prediction of abuse.

Racial/Cultural Bias in Risk Assessment

There are great variations across professional groups, cultural, ethnic, and religious groups, and by geographical location, as to what constitutes abuse (Collier et at., 1999; Gelles, 1982; Giovannoni & Beccera, 1979; Korbin, 1981; Starr, 1982; Valentine, et al., 1984). It is increasingly argued that child abuse in not only multi-faceted, but is socially constructed. It seems that racial/cultural bias influences public response to the

social problem of child abuse and neglect in the U.S. (Brissett-Chapman, 1997). A disproportionate number of African American families come to the attention of CPS despite survey research that suggests that the actual incidence of maltreatment within the general population is similar for African American and majority families (Sedlack & Braodhurst, 1996). African Americans are also more likely to have their children removed from the home subsequent to CPS investigation (Baird, Ereth & Wagner, 1999). It seems that African Americans are under increased surveillance in regard to abuse and neglect. Some have argued that official definitions of abuse and neglect have emerged without adequate input from minority voices, and that this bias extends to the systems employed to assess risk (Brissett-Chapman, 1997; Lyons, et al., 1999).

Although racial bias in child protective services is a topic that warrants much more attention, it should not be assumed that formal risk assessment contribute to this problem. To the contrary, structured risk assessment procedures, especially actuarially derived models, may offer a partial remedy to biased assessment and decision-making. Johnson (1994) found that an empirically-derived risk assessment used in Alameda County, California not only outperformed clinical predictions in overall accuracy of prediction of recurrence of maltreatment, but also was superior in predicting recurrences specific to African American, Caucasian, Hispanic, and Asian samples.

A study conducted by the Children's Research Center of the National Council on Crime and Delinquency examined relationships between risk, race, and recurrence of abuse and neglect in California, Georgia, and Michigan, states that have implemented actuarial risk assessment systems. In jurisdictions studied, essentially equal proportions

of Whites and African Americans were classified to each of four levels of risk. In Michigan, Whites, not African Americans, had a marginally higher risk rating when the actuarial assessment was applied. Not only did these systems classify similar proportions of cases of each racial/ethnic group to each risk level; each group exhibited similar rates of subsequent maltreatment within each risk level (Baird, Ereth, & Wagner, 1999).

While far from definitive, this initial research suggests that formalizing risk assessment procedures may reduce, rather than increase, racially biased judgements in CPS settings. After all, it must be remembered that caseworkers will make such assessments with or without scientifically derived tools. There is little reason to believe that unaided caseworker judgement would be less racially biased than structured risk assessment models.

Performance of Current Risk Assessment Systems

Risk assessment systems currently used by CPS agencies fall into two general categories, "consensus-based" and "actuarial" systems. The primary purpose of most versions of both types is to help caseworkers to identify the highest risk cases so that limited resources can be targeted toward them (Schene, 1996). What differentiates the two categories of systems lies in how they are developed. Consensus-based models are developed by committees of practitioners, administrators, and other experts, based on discussions and review of available literature (Johnson, 1996). Illinois is generally cited as having developed the first consensus-based CPS risk assessment model (Fluke, 1994). This model consists of 17 risk factors, and was developed by the process described

above. Other states soon followed, adopting or modifying the Illinois model. Other states began independently developing their own models through consensus (Baumann, et al., 1997). Some models, such as the Structured Model for the Assessment of Risk in Texas (SMART) are consensus-based models that have been revised through empirical study, and thus may be considered "blended" models, derived both by consensus and actuarial methods (Baumann, et al., 1997).

Actuarial risk assessment systems (also called empirically-based systems) are developed by examining samples of cases whose characteristics at the time of case opening and whose outcomes (e.g., maltreatment recurrence) are known when the study begins (Johnson, 1996). Only those factors that are demonstrated to correlate with selected outcomes are included in these models. The earliest systems of this type were developed in the mid 1980's in Alameda County, California by Johnson and L'Esperance and in Alaska by Baird and colleagues from the National Council on Crime and Delinquency (NCCD). Michigan soon followed and now has in place perhaps the most thoroughly researched model of this kind (Baird, 1999). The NCCD has continued to develop such models for various states and actuarial risk assessment systems are now being used in New Mexico, Vermont, Wisconsin, Rhode Island, and Oklahoma. As mentioned, Texas uses a "blended" model that is largely actuarial in nature, in that it has been refined through empirical methods. Studies of the reliability and validity of several versions of each type of risk assessment system have been conducted over the past two decades.

<u>Measuring Model Reliability and Validity</u>

<u>Measuring Reliability.</u> Nunnally defines reliability as, "the extent to which measurements are *repeatable*—by the same individual using different measures of the same attribute or by different persons using the same measure of an attribute"(1967, p.172). Risk assessment studies that have reported reliability data discuss *internal consistency, interrater reliability*, or both. Internal consistency refers to estimates of reliability based on the average correlation among items within a test or assessment tool (Nunnally, 1967). All risk assessment reliability studies reviewed use *Cronbach's alpha* as the reported measure of internal consistency. Interrater reliability is concerned with consistency across users of the system (Pecora, 1991), that is, to what extent different workers using the model agree on ratings assigned to cases.

<u>Measuring Validity.</u> Nunnally (1967) states that, "In a very general sense, a measuring instrument is valid if it does what it is intended to do" (p.75). A yardstick is a valid instrument insofar as it accurately measures distance between points. *Validation* is the process of establishing validity through empirical investigation. Validity is a matter of degree, rather than an all-or-none property. Several ways of determining the validity of a measure are available. For instance, *face validity* is a type of validity that simply refers to the degree to which a measure *appears* to measure an intended construct, or idea, through examination of its contents. The types of validity necessary to support the use of a given measure are mainly determined by the intended use of the instrument (Nunnally, 1967). *Construct validity* refers to the extent to which a measure adequately measures a "construct", something that does not exist as an isolated, observable dimension of

behavior (Nunnally, 1967). Construct validity is typically tested by determining the degree to which items which make up the instrument in question correlate (converge) with those of other *validated* measures of the same construct, or conversely, fail to correlate (diverge) with measures of unrelated constructs. Construct validity is especially important to establish when attempting to measure highly abstract concepts, such as anxiety.

However, in relation to the prediction of child physical abuse, the truly meaningful measure of a given system is its *predictive validity*. Predictive validity is at issue when the purpose is to use an instrument to estimate some form of behavior (criterion) (Nunnally, 1967). The "prediction" can be of an event that has already occurred (post-diction), a condition that exists simultaneously (concurrent validity), or a future event (prediction). Whichever the case, the logic is the same; it does not matter when the data are available (Nunnally, 1967). Reliability and predictive validity findings from studies of consensus-based and actuarial risk assessment systems are summarized below.

Consensus-based Systems Performance

[SEE TABLE 1 – CONSENSUS-BASED SYSTEMS PERFORMANCE]

Reliability. Reliability data on consensus-based systems have been mixed. Baird (1999) compared the interrater reliability of two consensus-bases risk assessment systems (Washington model, 'WRAM' and Illinois model derivative, 'CFAFA') and one actuarial system (Michigan Family Risk Assessment of Abuse and Neglect, 'FRAAN'). Eighty case files were reviewed and rated for risk by twelve case readers.

Table 1-Consensus Based Systems Performance

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
The Illinois Model (CANTS 17B)	Baird; Wagner; Healy; Johnson (1999)	Reliability (Compared to Michigan and Washington models)	12 Independent ratings of 80 case files	simple % agreement Cohen's Kappa correction	Generally weak reliability
	Calica, Cotton, Edwards(1998) Child Endangerment Risk Assessment Protocol (CERAP)	System effects	Total CPS Case recidivism pre and post RA implementation	simple % , survival analysis	Significant reduction in recurrence rate post implementation
	Fuller and Wells (1998)	System effects (Recurrence of abuse in population)	Case control Design Short term "indicated" recurrence	logistic regression	Lower recurrence where RA was completed
	Commasso and Jagannathan (1995)	Predictive validity (Case substantiation, Closing, Recidivism)	Case file abstraction review of 102 cases-7 coders	ROC Curve Analysis, logistic regression, Wilcoxin	Generally Poor Predictive validity
Philadelphia Model (CANTS derivative)	Fluke et al. (1993)	Interrater reliability System effects, Service effects	CARF; WARM; Philadelphia; retrospective multimedia	Cronbach's Alpha; chi square; bivariate analysis	Interrater-marginal; Internal consistency-good; System and service effects marginal

Table 1 *(continued)*

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
Washington Assessment of Risk Matrix (WARM or WRAM)	Baird, Wagner, Healy, Johnson (1999)	Reliability compared to Michigan and CANTS (derivative) models	12 Independent ratings of 80 case files	simple % agreement Cohen's Kappa correction	Marginal reliability
	Comasso and Jagananannathan (1995)	Predictive Validity (Case substantiation, Closing, Recidivism)	Data abstraction case review of 102 cases-7 trained coders	ROC Curve Analysis, logistic regression, Wilcoxin	Generally poor, but better than chance
	Fluke, et al. (1993)	Interrater reliability Internal consistency System effects Service effects	CARF; WARM; CANTS 17B (derivative): Comparative Analysis	Cronbach's alpha; Chi square; bivariate Analysis	Fair interrater-reliability; Good Internal consistency, Marginal system and service effects
	Marks and Mcdonald (1989)	Predicyive validity: Case Closing, Substantion, Recidivism	Case record review	discriminant analysis; logit analysis	Total correct =81% (DA), or 70.5% (Logit) fp=19.6% or 2.9% fn=14.3% or 85.7% (DA vs. Logit)

(Note: fp = false positive, fn = false negative)

Table 1 (*continued*)

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
Child at Risk Field	Koiko (1998)	Reliability; Convergent Validity	% agreement across Allegation, Caseworker decision and Individual informants	Cohen's Kappa; chi square; Pearson's r	Generally good reliability across sources; Poor convergence with clinical measures
	Doeck, English, Depanfilis, Moote (1993)	Process	Data abstraction	chi square; bivariate analysis; principal component analysis	Implementation problems; 4 underlying factors
	Fluke, et al. (1993)	Interrater reliability; Internal consistency; System effects; Service effects	CARF; WARM; CANTS 17B (derivative); Comparative analysis	Cronbach's Alpha; chi square; bivariate analysis	Fair reliability; Marginal system and service effects

44

Table 1 *(continued)*

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
Child Well Being Scales (CWBS)	Doueck & Lyons (1998)	Prediction of caseworker activity	133 item data abstraction n=337	hierarchical regression	Modest improvement in prediction of worker activity
	Faushel, Finch & Grundy (1994)	Internal consistency; Predictive validity (prediction of caseworker's perception of risk)	CWBS, CPSRD, Problems and Conditions form	multiple regression	Good internal consistency; Sig. correlation with workers perception of risk
	Nasuti & Pecora (1993) UTAH Model (CWBS derivative)	Interrater reliability	CPS "experts" n=15 and caseworkers n=56 rated 3 case vignettes	Cronbach's Alpha; Pearson's r; Spearman Brown Prophecy	Interrater rel=.75 Cronbach's a=.77 All: Cronbach's a>.94
	Gaudin, Polansky, & Kilpatrick (1992) (CWBS Segment)	Interrater reliability; Concurrent validity	Control, n=80 Neglect group, n=53	correlation; bivariate analysis; discriminant analysis	All: Cronbach's a>.97 fp=13%; fn=21% Rc = .72
Child Protective Services Review Document (CPSRD) (New York)	Fanshel, Finch & Grundy (1994)	Prediction of caseworker's perception of risk	CWBS, CPSRD, Problems and Conditions form-12 volunteer workers	multiple regression	Drug and alcohol segment correlated with workers perception

(Note: fp = false positive, fn = false negative)

Two measures of interrater reliability were calculated, simple percent agreement and Cohen's kappa correction. Cohen's kappa gives percent agreement, adjusted for chance, that is the percent agreement that would result in random assignment of ratings. A kappa of 1.0 indicates perfect agreement, whereas 0.0 indicates agreement equal to chance. A negative kappa would indicate agreement worse than chance (Orme, 1986).

Results were not impressive for the consensus-based models. Raters using CFAFA agreed on only 13 of 80 (16.3%) of cases and those using the WRAM agreed on only 11 of 80 (13.8%). The kappa corrections for the CFAFA and WRAM were .184 and .180 respectively, well below the generally accepted threshold of .5 to .6. Both models produced interrater reliability ratings that were inferior to those produced by the actuarial model.

Consensus-based models have fared better in some research. Nasuti and Pecora (1993) estimated overall interrater reliability scores above .97 using the Spearman-Brown prophecy formula for 22-28 raters, who rated 3 case vignettes. However these results may be inflated due to the fact that the correlation coefficients obtained were summated scores that best represent an average score for each vignette. The results do not provide evidence as to whether the ratings were made in the expected direction. Also, the small number of vignettes viewed may not have provided an adequate sample to approximate the conditions that may be encountered in the field. Fluke et al. (1993) evaluated the reliability of three consensus-based models, WRAM, CARF, and Philadelphia Model (Illinois CANTS derivative) by examining 25-50 caseworker ratings of two case vignettes. Using Cronbach's alpha coefficients, these authors found good reliability for

group mean correlations (above .94 for each model), but poor *single-rater* reliability (.25

to .60). These coefficients represent the upper and lower bounds of expected reliability,

respectively. Again, the very small number of case vignettes used may have produced an

overly controlled experimental condition with limited generalizability to actual CPS

practice. Overall, these results reflect generally poor interrater reliability for these

models.

Several consensus-based models have been shown to have high levels of internal

consistency. Nasuti and Pecora (1993) reported a Cronbach's alpha of .77 f or the Child

Well-Being Scales (CWBS), developed by the Child Welfare League of America

(Magura, Moses, & Jones, 1987). The CWBS consists of 43 behavior-rating scales that

are multidimensional measures of potential child maltreatment situations. The items

represent four dimensions linked to child well being: parenting role performance, familial

capacities, child role performance, and child capacities. Each item consists of four or five

response categories with anchoring points. In addition, each scale point is weighted based

on the "seriousness" of the condition (Lyons, et al., 1999).

Gaudin, Polansky, and Kilpatrick (1992) report a group-mean alpha of .94 for a

selected segment of the CWBS. Fanshel, Finch, and Grundy (1994) also found good

internal consistency for the CWBS, as well as for New York City's Child Protective

Services Review Document (CPSRD).

Overall, consensus-based risk assessment systems have demonstrated good

internal consistency; meaning that the individual items contained in the scales are highly

correlated with one another. However, tests of interrater reliability have yielded mostly

poor results, indicating that workers' ratings on these forms vary widely. This suggests that different caseworkers using these tools in actual practice are likely to vary in their assessments of risk for a given family.

Predictive Validity. Support for the ability of consensus-based risk assessment systems to predict recurrence of abuse is generally weak. Camasso and Jagannathan (1995) conducted the most scientifically rigorous study of two consensus-based models, WARM and Illinois CANTS 17B. The authors introduced Receiver Operating Characteristic (ROC) curve analysis to the evaluation of risk assessment models. This procedure allows for comparison of specificity and sensitivity across various diagnostic cut points, compensating for the arbitrariness of a set cut point. In their analysis, the authors found the models to predict case closings, case substantiation, and recidivism better than chance, but not very well overall. The WARM was somewhat superior to the CANTS 17B in the prediction of recidivism, with the most variance explained by the WARM's severity and child behavior problem scales.

Marks and McDonald (1989) evaluated the predictive ability of the WARM using discriminant analysis and logit analysis. Discrimanant analysis produced superior results in prediction of overall rates of recurrence, 81.0 %, versus 70.5 %. Discriminant analysis also produced higher *sensitivity*, correctly predicting 86 % of recidivists, versus only 14% with logit regression, while logit regression produced better *specificity*. However, Comasso and Jagannathan (1995) criticized the methodology used in this study, stating that the variations were mainly an artifact of different arbitrary cut points generated by the statistical software packages for each model.

Actuarial Systems Performance

[SEE TABLE 2 – ACTUARIAL SYSTEMS PERFORMANCE]

Reliability. Results from research done thus far have shown actuarial risk

assessment systems to produce more reliable ratings than consensus-based models. Baird

et al. (1999) found interrater reliability to be better for Michigan model (FRAAN), as

compared to the Washington (WRAM) or Illinois derivative (CFAFA). In 85% of all

cases (n = 80) rated by twelve case readers, as least three of four raters assigned the same

risk level. This level of agreement was achieved on 51% of cases using the WRAM and

45% using the CFAFA. The FRAAN produced a kappa (computed median value for all

sets of raters) of .562, significantly better than the WRAM (.180) and CFAFA (.184). A

kappa of .5 to .6 is generally deemed acceptable for research purposes, and 1.0 would

indicate perfect agreement (Baird et al., 1999).

Baumann et al. (1997) assessed the interrater reliability and internal consistency

of the SMART (Structured Model for the Assessment of Risk in Texas), the risk

assessment component of the broader WISDOM (Worker Improvements to the Structured

Decision and Outcome Model) developed and implemented by the Texas Department of

Protective and Regulatory Services as a decision enhancement project. This is a

"blended" model, in that the original consensus-based system has been modified through

empirical study. 110 caseworkers viewed three video taped fictitious case scenarios

(abuse and neglect) and assigned risk ratings. Interrater reliability was exceptionally high

(.996). Internal consistency of the model was weak. The alpha coefficients for three

subscales of the instrument fell below the minimum acceptable level of .70. The authors

Table 2-Actuarial Systems Performance

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
New Mexico Model	Meyer & Wagner (1998)	(Classification: into levels of risk)	Re-abuse investigation and substantiation rates within 18 months- Both on those originally substantiated and unsubstantiated	simple % in each risk category	Good classification rates achieved
Alaska Model	Baird(1988)	Predictive validity	Case abstraction	correlation; multiple linear regression; crosstabs; chi-square; Pearson's r	Abuse scale: fp=21.9%, fn=31% Total correct=76%
Alameda County	Johnson & L'Esperance (1984)	Predictive validity (physical re-abuse)	105 item data abstraction completed by 10 graduate assistants	multiple linear discriminate analysis	fp=18%, fn=30.4% Total correct=74.4%

(Note: fp = false positive, fn = false negative)

Table 2 *(continued)*

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
Michigan Model	Baird et al.(1999)	Interrater reliability	Comparisons with WARM, CFAFA(CANTS Derivative); 12 readers-80 cases	simple % agreement; Cohen's Kappa correction	Significantly better reliability over other 2 models
	Baird et al.(1995)	Service effects over 12 month period	13 pilot counties matched with cohort	simple comparisons of rates recurrence	Fewer new substantiated referrals and placements in pilot counties
Wisconsin Model	Neuenfeldt & DeMares (1994)	Classification accuracy-4 levels of risk	Re-abuse allegations, substantiations, injuries; over 12-24 months	simple % in each categories	Actuarial model performed well, similar to results achieved in other states
Vermont Model	Weedon, Torti, Zunder(1998)	Predictive validity	Administration of Vermont Model to open cases	regression, discriminant analysis	fp=14%; fn=62% Total Correct=68%

(Note: fp = false positive, fn = false negative)

Table 2 (continued)

Model	Authors and Date	Information Available	Measures	Statistical Analysis	Results
Oklahoma	NCCD (2000)	Classification Accuracy across 5 risk levels	Subsequent substantiated abuse or neglect 18 month follow-up	simple % correct	Highest risk group 15 times more likely to re-abuse than lowest risk group
Texas (Wisdom) Blended-Consensus/Actuarial	Baumann et al. (1997)	Internal Consistency; Interrater reliability; System effects	Randomly selected caseworkers (n=102) and (n=110) rated 3 videotaped vignettes recurrence of maltreatment pre and post RA implementation	average interrater correlations; Cronbach's Alpha; survival analysis	Good interrater reliability; Fair to poor internal consistency; Lower recurrence post-implementation
Rhode Island	Squadrito & Wagner (1993)	Classification accuracy-4 risk levels	Subsequent substantiated abuse; Subsequent hospitalization	simple % Correct	Good classification for both recurrence and severity

concede that the interrater reliability estimates are artificially inflated due to missing information in the case presentation scenarios, resulting in many "unknown" answers. Also, as discussed regarding evaluation of consensus-based models, such a small number of case scenarios undermines the credibility of the study to fairly replicate the actual CPS practice (Baird, et al., 1999).

Actuarial models tend to produce higher levels of interrater reliability than do consensus-based models. Actuarial models tend to be shorter, and to use case characteristics that flow mechanically from case files (Johnson, 1996). Therefore, these systems may require less work and intuitive judgement to complete, enhancing consistency of ratings across workers completing them.

Predictive Validity. The performance of actuarial systems in the prediction of recurrence of abuse can best be described as marginal. Johnson and L'Esperance (1984) randomly selected 120 cases of physical abuse from case files in Alameda County, California. Using eighty-one of these cases, the authors developed an actuarial predictive model by examining correlations between case characteristics and recurrence of abuse (defined as re-referral). The remaining thirty-nine cases were used as a cross-validation sample. Multiple linear discriminant analysis produced an overall correct classification rate of 74%, with 18% false positive and 30% false negative predictions. A particular strength of this study is in its use of statistical controls of treatment effects. Three control variables were constructed to control for duration and intensity of CPS contact.

An evaluation of the Alaska model produced similar results. Overall classification accuracy was 76%, with 22% false positives and 31% false negatives (Baird, 1988). A

53

study of the Vermont model showed it to correctly classify re-abuse and re-neglect in

68% of cases. However, false negative predictions were twice as high as in the other

studies (62 %), while false positives were similar at 14% (Weedon, Torti, & Zunder,

1988).

Recently, some researchers have argued that *prediction* of abuse is not the

appropriate measure of actuarial risk assessment systems. Instead, *classification*, or

grouping of certain types of cases as more likely than other types to result in subsequent

maltreatment, is a preferable strategy (Baird, et al. 1999; Schene, 1996). In this approach,

"high-risk" simply denotes inclusion in a group with significantly higher rates of

recidivism; it is not a prediction of failure. Baird and colleagues state that other fields in

which low base rate phenomena are of interest, such as medicine, juvenile justice, and

adult corrections have abandoned the idea that risk assessment is an exercise in prediction

(1999). This represents a somewhat less ambitious goal than do attempts to predict an

outcome for a specific case: "Prediction" is more precise than "classification". It is also

argued that "classification" has more utility in regard to decisions that CPS workers must

make. "Prediction" implies that dichotomous decisions will be made. Whereas, risk

assessment in CPS is mainly intended to provide a continuum of risk levels in order to

guide a range of decision-making responses (Baird, et al., 1999).

Most recent research on actuarial risk assessment systems report *classification*

data, versus predictive validity. Applying this standard, current actuarial systems (many

developed for CPS by the NCCD) perform fairly well. New Mexico conducted an

evaluation of its actuarial system with the assistance of the Children's Research Center

(CRC) of the NCCD. The model was found to accurately classify cases across four levels of risk of recurrence of maltreatment over a period of eighteen months (Meyer & Wagner, 1998). Two outcome variables were defined (recurrence of investigation and recurrence of substantiated maltreatment). The model performed well in classification of cases that were initially substantiated, as well as for those originally unsubstantiated. This demonstrates that this system is also potentially useful for targeting of services to those families that are initially *investigated*, but not *confirmed* for maltreatment. The authors suggest that *risk* is a better criterion for service provision than is *substantiation status*.

Actuarial systems employed in other states have performed similarly. For instance, in Oklahoma, the CRC-developed model classified cases into five levels of risk. Cases in the highest risk category re-abused at a rate of 57%, whereas those in the lowest category re-abused at a rate of 3.7% (NCCD, 2000). Rhode Island's system was found to accurately classify cases in regard to recurrence and severity (Squadrito & Wagner, 1993). Severity was defined as risk of suffering subsequent abuse requiring a hospital visit.

It seems that some researchers and users of actuarial risk assessment systems have concluded that classification of cases along a continuum of categories of risk is more feasible, and maybe more useful, than attempting to predict recurrence of abuse for a specific case. The actuarial models tested thus far do appear to perform reasonably well at this task. It is difficult to make comparisons with consensus-based systems using this standard because research has not been directed toward determining the *classification accuracy* of consensus-based systems.

Statistical Analyses Used in Risk Assessment Research

The predictive validity studies have typically used discriminant analysis and/or multiple regression to analyze data. In a critique of risk assessment systems, Lyons, et al. (1996) state that these methods may be inappropriate with low-base-rate phenomena, small sample sizes, and dichotomous variables, because the assumptions of normality and equal covariance are likely to have been violated (p. 149). Also, accounting for interactions among predictor variables can be problematic when using multivariate statistics. Interactions among predictor variables must be anticipated by the researcher and modeled a priori. Adding interaction variables requires increasing sample size in order to maintain statistical power (March, 1998). The typical rule of thumb when using multivariate statistics is that at least ten subjects are required for each predictor variable (Cohen, 1977). As noted above, several influential interactions are known to be associated with the prediction of physical child abuse recurrence. Other, unidentified interactions may exist. Regression methods that rely on examining covariance in a pairwise fashion tend to over-select explanatory variables from a larger set of possible predictors, emphasize main effects at the expense of interactions, and often pass over causally related variables for secondary or redundant ones (Marshall & English, 2000). Adding interaction terms necessitates a larger sample size to maintain statistical power.

Recent studies of child abuse recurrence have applied survival analytic techniques (DePanfilis & Zuravin, 1999; Fluke, et al., 1999). These techniques allow for the inclusion of *time* as an outcome measure. The dependent variable is defined as time to recurrence, versus recurrence as simply a dichotomous (yes or no) variable. These

techniques have not yet become popular for the testing of models intended to predict the recurrence of physical child abuse. However, in a related study, Bolen (1998) compared models of risk of sexual abuse derived from survival analysis and logistic regression. The models performed similarly in regard to bivariate statistical outcomes but the survival analysis allowed for visual identification of interactions between age at time of abuse and other predictor variables.

Event history techniques, such as survival analysis, may offer a more realistic measure of outcomes for families who enter CPS services. It is informative to know not only which families are a higher risk to re-abuse, but also which are likely to "survive" (experience no new abuse) over a period of time. This could be particularly useful in the evaluation of different treatment services. For instance, a particular treatment may be considered successful, to some degree, if it were to increase time between abusive episodes. However, event history techniques have been described as being especially problematic in the examination of low-base-rate events (Fluke, et al., 1999).

Receiver operating characteristic (ROC) curve analysis, introduced to the risk assessment literature by Comasso and Jagannathan (1995) is especially useful because it allows for comparison of specificity and sensitivity across various diagnostic cut points. Users of risk assessment instruments may use this analysis to maximize the sensitivity or specificity, as desired, based on the intended use of the instrument. The typical risk assessment analysis attempts to predict the likelihood of an outcome event, such as case substantiation or recidivism, using a binary outcome estimation procedure such as discriminant analysis, probit, or logistic regression. Probability levels (cut points) are

imposed knowingly or by default (e.g. .50) with predictions greater than chance being classified as "events" and those less than chance as "nonevents". The model's sensitivity and specificity are greatly influenced by the selected probability level.

ROC curve analysis provides a means for assessing the discriminating ability of an instrument across a spectrum of diagnostic "cutpoints". Each cutoff point is plotted with respect to the "true positive" and "false positive" rates that result from the criterion selection (Comasso & Jagannathan, 1995). This method allows the researcher to examine changes in sensitivity and specificity at various cutpoints. Interestingly, ROC curve analysis has not become widely used for model evaluation since this study. This may be due to the recent trend in tests of actuarial models to opt for inspection of simple classification rates across a small number of categories of risk (low, medium, high), instead of *prediction*, per se.

In addition to expert knowledge and conventional statistical methods, researchers have begun to explore the power of modern computer programs as potentially useful tools to facilitate CPS decision-making. The Seventh National Roundtable on CPS Risk Assessment featured a small group discussion to examine the potential of two types of "artificial intelligence" programs, "expert systems" and "artificial neural networks", for the assessment of risk of child abuse. While some reservations were noted, the group generally felt that the technology was promising and deserved further research (Fluke, 1993). While little research has followed, these technologies continue to receive mention within the professional literature. In a 1996 article to the American Professional Society on the Abuse of Children (APSAC), Sheets states,

As long as risk models remain document-based, fixed, as it were, in a two-dimensional medium, data fragmentation will be a significant problem. Words on paper cannot talk back to us in a dynamic way. But computers can. With their ability to present a depth and breadth of information in a holistic, three-dimensional, multimedia format and their ability to "think" and even to "learn" from new information, computers show promise of providing a decision-support technology that fits the CPS decision making environment well and acts as a useful partner for those beleaguered decision makers, us. (p.11)

The Potential of Artificial Intelligence Technology to Improve Risk Assessment

The term artificial intelligence (AI) describes a variety of computer applications that are designed to model certain aspects of human intelligence and to acquire knowledge to solve intellectually challenging problems (Schoech, 1999). These systems have evolved over the last forty years. Early systems attempted to model complex cognitive processes such as reasoning, thinking, learning, and creating. The goal was to develop general-purpose problem-solving machines. This proved to be a daunting task and the most current efforts are aimed toward more delimited domains of cognitive functioning. The most predominant AI technology in human service literature is "expert systems" (Schoech, 1999). These systems are computer programs that embody the knowledge of human experts in order to consult on real-world problems (Gingerich, 1990). Since expert systems employ knowledge imparted by human experts, a risk assessment system embedded in such software may be considered a special type of

consensus-based system. Artificial neural networks (ANNs, or NNs) represent another type of AI application that has gained recent attention in social work research. Neural network applications use data analysis models based on human brain analogies of neurons (nodes) and synapses for classification and prediction problems (Schoech, 1999). These two types of computer-based models may prove useful to support decision-making in CPS. Their potential and limitations will be examined below. Occasionally, the term "artificial intelligence" is used somewhat interchangeably with the term "expert systems". In this paper, AI is used to describe the broader range of applications that encompasses both expert systems and artificial neural networks.

Expert Systems

Expert systems have been used in human services for diagnosis and treatment of infection (Buchanan & Shortliffe, 1984), therapy advice for depressed patients (Mulsant & Servan-Schreiber, 1984), classifying learning disabled students (Ferrara, Parry & Lubke, 1985), recommending behavioral intervention strategies in the classroom (Serna, Baer, and Ferrara, 1986), and advising on treatment of emotional crisis in remote areas (Hedlund, Viewing, & Cho, 1987). In social work they are being developed to advise on child placement decisions (Schuerman, 1987), intervention design in family therapy (Goodman, Gingerich, & de Shazer, 1989), and services for battered women (Nurius & Nicoll, 1988). However, little research has been conducted to evaluate the predictive or classification accuracy of these models.

Few human service expert systems are currently in use. One, called ECS (expert counseling system) is being used to identify employment problems, set goals to resolve these problems, provide tips on how to do this, and make referrals to appropriate resources. Another expert system is being used to check patients' medication orders against hospital records to reduce adverse drug interactions. This system reportedly prevents 1.2 drug-related mistakes each day (Schoech, 1999).

Martindale, Ferrara, and Campbell (1987) evaluated the performance of an expert system that classifies students as learning disabled. The expert system was found to be in general agreement with the recommendations of a panel of human experts. Schuerman and Vogel (1986) reported developing an expert system to support placement planning decisions in child welfare. However, the validity of conclusions produced by the system has not been reported.

An expert system asks a series of questions, then applies *rules* in its knowledge base to deduce the nature of the problem and arrive at a recommendation. These are called "if-then" rules; "Ifs" refer to antecedents, "thens" to conclusions. Systems vary greatly in complexity. Some take months or years to *train* (Gingerich, 1990). Training consists of defining the if-then rules, based on human expertise.

Expert systems rely on knowledge acquired from human experts. Knowledge is obtained through consultation of experts in a certain domain. Obtaining this knowledge can be difficult because it may be difficult for experts to describe their reasoning in a step-by-step fashion. In developing an expert system, it is important to establish an

acceptable level of performance by which to validate the application. Generally, the expert system should produce results equal to those of human experts (Gingerich, 1990).

Expert systems may prove useful in making scarce expertise widely available. One study estimated that it takes two years for a worker to become proficient in CPS work. Yet, the average tenure of workers is only one and a half years (Gingerich, 1990). Expert systems can be used for training and tutoring and mining and refining practice wisdom. As "rules of thumb" that are used in practice become more elaborated, the theory becomes more refined. A particular strength of expert systems is in their transparency. A good system not only reaches conclusions, but also can tell the user by what means the decision was reached. They can serve to make decision factors explicit (Schuerman & Vogel, 1986).

Expert systems have obvious limitations. They can be costly to develop and train. They are useful only in areas where problems can be solved using available technology, generally a problem or task that could be performed by an expert in 10 to 30 minutes (Gingerich, 1990). Schuerman (1987) projected that the most likely candidates for use in CPS initially is in circumscribed decisions such as child placement or evaluating the risk of child abuse. While these systems may be useful in guiding placement decisions at some point, it does not seem that they will be particularly useful for predicting abuse. Expert systems require human expertise. Such expertise does not yet exist for the prediction of child abuse. Baird and colleagues (1999) found wide disparity in decision-making among caseworkers *and* among identified CPS "experts".

Another limitation of expert systems is that the "rule base" within the system might contain subjective biases as well as valid expertise about the problem (Gingerich, 1990). Like textbooks and training manuals, these systems are simply ways of storing and accessing human judgements, and thus are subject to any biases of the contributing human experts.

While expert systems have been touted as potentially useful tools to provide consultation to caseworkers for some time (Schuerman, 1987; Schuerman & Vogel, 1986), they have not become widely used or studied in CPS settings. Johnson (1996) concluded, "We are aware of no studies demonstrating that expert systems produce consistent and accurate ratings" (p.17). To date, there is no published research examining the usefulness of expert systems for the prediction of child abuse.

Artificial neural networks

Artificial neural networks (ANNs) are computer-based learning systems that have demonstrated utility in prediction, classification, and decision-making (Cross, Harrison, & Kennedy, 1995; Galletly, Clark, & McFarlane, 1996; Patterson & Cloud, 2000; Zou, Shen, Shu, Wang, Feng, Xu, Qu, Song, Zhong, Wang & Lieu, 1996). Unlike Expert systems, ANNs do not require highly specialized human expertise. They do not need a priori specification of decision paths, but instead learn to detect relationships between variables in decision sets. These applications may be useful where the underlying relationships among predictor variables are less well known compared with highly structured expert systems or equation-based approaches (Garson, 1998).

ANNs identify patterns between input (predictor) and target (criterion) variables. The terms "neurocomputers", "artificial neural systems" (ANS), and artificial intelligence are often used synonymously with artificial neural networks. These systems are typically described using the biological analogy of brain neurons. This analogy centers on the fact that neural networks do not operate on a set of programmed instructions, as do statistical packages. Rather, they pass data through a multiple parallel processing entities (nodes or neurodes) that "learn" and adapt to the patterns that are presented to them. Data are not stored in these entities, nor are particular answers stored at particular addresses (Garson, 1998). Processing functions assume a pattern throughout the system. This pattern, developed in the iterative learning process, comes to represent the relationships between the input variables and the target variable (Patterson, 2000).

ANN Architecture

The three basic components of a neural network are the neuron itself, often called a node, the "interconnection typology", consisting of the weighted paths between nodes (analogous to neural synapses), and the learning scheme. The task of the learning scheme is to optimize classification accuracy in "pattern recognition", which is generally defined as a process of discovery of meaningful relationships between explanatory variables and a response (output) variable (Cloud, 1999; Collins & Clark, 1993). The most commonly used learning scheme is called "backpropogation". In this procedure, "error terms", defined by differences between observed and desired outcomes, are fed back through the system in order to readjust parameter weights. Backpropagation neural networks "learn"

to classify a pattern through induction, by repeatedly processing examples of each class of the target variable (Gordon, 1992).

The most widely used and empirically tested type of ANN is the multi-layer perceptron (MLP), a type of backpropagation neural network model. These are descendents of simple perceptrons, devices that linearly discriminate whether an input pattern belongs in one of two categories. Simple perceptrons have two "layers" of nodes, one for predictor (independent) variables, and one for target (dependent) variables (Garson, 1998; Marshall & English, 2000; Patterson & Cloud, 2000). In addition to input and output layers, MLPs have one or more "hidden layers" of nodes between the input and output layer. Knowledge is represented in the weights assigned to the connections between nodes on different layers. These weights are initially set to small random values in the range (-0.1, +0.1). The weights are then repeatedly adjusted as input/output pairs are fed through the model (Gordon, 1992).

"Learning" takes place as errors (discrepancies between inputs and targets) are backpropagated (fed backward) through the network in a recursive fashion. The magnitude of error in classifying the input is used to determine the amount of change needed in each weight in order to reduce the error for the next presentation of the input. An error signal is only back through the network when the discrepancy between the network's output and the actual classification exceeds the "error tolerance", a threshold that is preset by the researcher (Heckert, 1994).

The required size of the training set depends on the number of input variables, the number of decisions to be made, and the complexity of the problem. With regression

methods we can calculate the size of sample needed to achieve a given level of confidence in the results. No such calculation is yet available for neural networks (Cross, et al., 1995).

While neural networks employ several preexisting concepts from the statistical literature, it is the combination of these that is novel (White, 1989). The net input to a given hidden unit is a discriminant linear function which, when subjected to a nonlinear transformation within the hidden unit, acts as a nonlinear feature detector. The output of all nodes in the hidden layer are then inputs to another linear discriminant function and another nonlinear transformation at each unit in the output layer (Gordon, 1992, White, 1989).

Potential Advantages of the ANN Prediction Approach

Neural networks offer some advantages over traditional statistical prediction methods. First, instead of assuming a particular form of relationship between independent and dependent variables, then using a fitting procedure to adjust the size of parameters in the model, neural networks construct a unique mathematical relationship for a given data set based on observed patterns between explanatory variables and designated outcomes (Marshall & English, 2000). The hidden layer of neurons and associated interconnections has no counterpart in discriminant analysis or multiple regression. The hidden layer of computational units or nodes, along with the associated learning algorithm, allows MLPs to find nonlinear solutions to complex problems (Gordon, 1992; Patterson & Cloud, 2000).

Gallinari, Thiria, Badran, and Goglman-Soulie (1991) analyzed the relations between discriminant analysis and neural networks while increasing the degree of nonlinearity between variables. The results showed an advantage for neural networks over discriminant analysis models that increased in magnitude as the nonlinearity of the problem increased. The authors demonstrated that each layer of weights in a network performs a nonlinear discriminant analysis from the states obtained in the previous layer. Each layer increases the separation and clustering of the different inputs and the last layer classifies the final projection (p.357).

Second, the nonparametric nature of neural networks may make them particularly suited to social science data, where normality and linearity cannot be assumed. Unlike regression models, neural networks do not assume the absence of interactions among input variables. ANNs have demonstrated capacity in handling interactions and nonlinearities (Garson, 1998). As discussed, interaction effects have been found among factors predictive of risk of child maltreatment (Zuravin, et al., 1994, Rosenthal, 1988). Whereas regression analysis requires the researcher to enter suspected interaction terms a priori, neural networks can fit linear, polynomial, and interactive terms without requiring the researcher to model them (Garson, 1998; Marshall & English 2000).

Third, neural networks have been demonstrated to be fairly robust in regard to handling input corrupted by random error (Hartzberg, Stanley, & Lawrence, 1990; Lippman, 1987; Weiss & Kurlikowski, 1991). Neural networks are not greatly influenced by any single input variable, but rather depend on a pattern of inputs. Neural networks degrade gracefully because information is distributed throughout the network. That is,

they maintain classification accuracy relatively well in the presence of random error and are less affected by problems like outliers (Garson, 1998).

Fourth, There is some evidence that suggests that ANNs may excel over linear discriminant models with increasingly stringent thresholds for class membership. A "threshold" refers to a minimum score that must be reached for an example to be classified into one of two classes (Gordon, 1992). Using simulated data, Gordon (1991a) found that a neural network and a discriminant analysis model performed comparably in classifying violent and nonviolent cases at decision thresholds of .50 and .70. But, beyond a .80 classification threshold the neural network maintained its classification accuracy, while discriminant analysis fell off at a steep decline.

The choice regarding the threshold beyond which positive diagnosis or conclusion is made can be assisted by the receiver operating characteristic (ROC) curve. ROCs allow the user to assess the effects of differing thresholds on the accuracy, sensitivity, and specificity of the system (Cross, et al., 1995). As will be seen below, much of the medical research comparing ANNs to other statistical methods reports differences between ROC curves.

Fifth, ANNs may perform well on problems with low base rates. Gordon (1991b) compared the accuracy of a backpropagation network and discriminant analysis on problems with decreasing base rates by using fictional data. Neural networks were found to perform similarly in classifying cases when base rates were .50 and .20. The discriminant analysis models performed poorly in classifying the lower base rate data. However, in another Monte Carlo study comparing neural networks to discriminant

analysis across various base rates, Heckert (1994) found no significant differences in classification accuracy. Further research is needed to determine if ANNs can improve prediction of rarely occurring phenomena.

Finally, it appears that there is no significant disadvantage, other than length of training time, in including a large number of predictor variables in neural network analysis (Hartzberg, Stanley, & Lawrence, 1990). The network will disregard variables that are not associated with the output by not assigning weights to those variables, leaving them at their near zero initial values. Also, it appears that intercorrelation among predictor variables does not detract from goodness of fit (Gordon, 1992). This would suggest that neural networks are suitable in classification problems in which the number of predictor variables is large, and the intercorrelation among those variables is high.

Neural networks have been purported to excel at prediction and classification problems, particularly when there are a large number of inputs that are related in nonlinear ways (Garson, 1998). In prediction, neural networks are an alternative to multiple regression, logistic regression, structural equation models, and expert systems. Neural networks may outperform traditional statistical procedures where problems lack discernable structure, data are incomplete, and many competing inputs are related in nonlinear ways (Garson, 1998).

ANN Performance

While relatively new to social science research, neural networks have been used more extensively in other fields, such as economics, medicine, and education (Garson, 1998). For example, Everson (1994) compared a backpropagation neural network with

multiple linear regression and discriminant analysis in a study of classification of educational performance. The neural network achieved higher classification accuracy, especially when underlying models were nonlinear. Hiemstra (1996) compared a neural network model with linear regression for the purpose of stock market prediction. The neural network outperformed the linear regression model by a wide margin.

ANNs have become increasingly used for prediction and classification of medical outcomes. A review of the medical literature reveals ANN applications in several specialties including cardiology, pediatrics, oncology, radiology, and psychiatry. Presented below is a review of selected medical research that has used artificial neural network modeling. Studies were chosen that used ANNs to predict negative medical outcomes (i.e. morbidity and mortality), and particularly those that have compared ANN results to other statistical and diagnostic methods. Findings from these studies are pertinent to efforts to predict child abuse recurrence, in that numerous, potentially intercorrelated independent variables are modeled in order to predict an undesirable outcome. Child physical abuse can be considered such an undesirable outcome, analogous to medical morbidity.

Dybowski and colleagues (1996) compared neural network analysis with logistic regression in prediction of mortality among a sample of 258 randomly selected critically ill patients. A comparison of ROC curves demonstrated the neural network to produce superior predictive accuracy. The authors concluded that the flexibility of the neural network in accommodating interactions among the predictor variables accounted for the superior performance. Pofahl and colleagues (1998) successfully applied a neural

network to predict length of hospital stay for acute pancreatitis patients using admission data from the patients' records. The ANN was able to predict with 75% sensitivity which patients would require more than seven days of hospitalization.

In cardiology, ANNs have been used to identify acute myocardial infarction (AMI). Baxt and Skora (1996) compared the diagnostic accuracy of a neural network to that of emergency department physicians in diagnosing AMI in 1070 patients presenting to the emergency department with a complaint of anterior chest pain. The ANN analyzed patient data that was gathered by the physicians. Only7 % of these patients actually had an AMI (a low base rate). The physicians' diagnostic sensitivity was 73% and specificity was 81%. The ANN produced a sensitivity of 96% and a specificity of 96%. The authors concluded that this type of non-linear computational analysis offers promise for assisting physicians in diagnosing AMI. In another study, a neural network, trained on 39 items of clinical and ECG data, provided better diagnosis of AMI than either discriminant analysis or a common serum myoglobin measurement. The authors recommended a combination of clinical opinion, ANN output, and myoglobin test results to improve diagnostic accuracy (Kennedy, Harrison, Burton, Fraser, Hamer, MacArthur, McAllum, & Steedman, 1997).

In pediatrics, ANNs have been used to predict mortality and intercranial hemorrhage in preterm neonates. In a study of preterm infants (N = 890), a neural network and a logistic regression model were developed to predict mortality. An analysis of ROC curves showed the ANN to outperform the regression model by a significant margin (Zernikow, Holtmannspoetter, Michel, Pielemeier, Hornschuh, Westermann, &

Hennecke, 1998). Another comparison of ANNs with logistic regression, this time using

intercranial hemorrhage in preterm neonates as the target outcome, also showed the ANN

to be significantly superior to the regression model (Zernikow, Holtmannspoetter,

Michel, Theilhaber, Pielemeier, & Hennecke, 1998). In addition to these mortality and

morbidity applications, ANNs have been used to estimate fetal weight. A neural network

was used to estimate fetal weight for 100 patients with suspected macroscopic fetuses.

The ANN produced and error rate of 4.7%, significantly better than the rate of 10%

obtained from regression models (Farmer, Medearis, Hirata, & Platt, 1992).

ANNs have proved useful for improving cancer diagnoses. Neural networks have

been successfully used to distinguish between malignant and benign pelvic tumors,

outperforming logistic regression (Timmerman, Verrelst, Bourne, De Moor, Collins,

Vergote, & Vandewalle, 1999) and serum assay (CA 125) tests (Zhang, Barnhill, Zhang,

Xu, Yu, Jacobs, Woolas, Berchuck, Madyastha, & Bast, 1999). Researchers in London

developed a neural network that, when trained on patient demographics and previously

gathered medical information, distinguished between benign (n = 52) and malignant (n =

15) ovarian tumors with 100 % sensitivity and 98 % specificity for preoperative patients

(Tailor, Jurkovic, Bourne, Collins, & Campbell, 1999). The researchers concluded that

ANNs may be useful in the prediction of ovarian malignancy, and recommended

prospective evaluation of their application in this arena.

Zhao and colleagues (1998) entered urinalysis data into a neural network in an

attempt to distinguish between cancer patients with 14 different kinds of cancers (n = 25)

and healthy patients (n = 25). The ANN was able to correctly classify 85% of test set,

significantly better than principal component analysis or canonical discriminant analysis (Zhao, Xu, Yue, Liebich, & Zhang). Bryce and colleagues (1998) used an ANN to predict survival of patients with advanced carcinoma of the head and neck. Examination of the ROC curve showed that the ANN outperformed a logistic regression model across various degrees of sensitivity and specificity and was able to model variables that could not be included in logistic regression (Bryce, Dewhirst, Floyd, Hars, & Brizel).

In radiology, ANNs have been used to improve diagnostic accuracy. Ashizawa and colleagues (1999) developed a neural network to assist radiologists in differentiation between eleven different pulmonary diseases. The researchers found that radiologists using the ANN data were able to differentiate among various diseases significantly better than those relying on visual examination of radiographs alone (Ashizawa, MacMahon, Ishida, Nakamura, Vyborny, Datsuragawa, & Doi). The researchers concluded that the neural network can provide a useful "second opinion" to assist radiologists in making differential diagnosis of lung disease. In another study, an ANN using radiographic and clinical information to predict active pulmonary tuberculosis was compared to clinicians' assessments based on the input data. The ANN was more accurate than clinicians in accurately identifying patients with contagious active TB (El-Solh, Hsiao, Goodnugh, Serghani, & Grant, 1999).

In psychiatry and neurology, ANNs have been used for diagnosis and outcome prediction. Zou and colleagues (1996) found that a backpropagation artificial neural network was superior to two other computer-based diagnostic programs in diagnosing psychiatric disorders. Jefferson and colleagues (1998) combined a neural network with a

genetic algorithm to predict depression after mania. Genetic algorithms are search procedures that can be used to assist in variable selection for ANN models. The "evolved" neural network outperformed a logistic regression model in prediction of mania in 100 subjects. A fully-connected ANN, not using the genetic algorithm, also outperformed the logistic regression model, but was inferior to the "evolved network" (Jefferson, Pendleton, Lucas, Lucas, & Horan). In another study, researchers entered eye-tracking performance measures into a backpropagation neural network to distinguish between schizophrenic and normal patients. The ANN correctly classified 80% of patients based solely on the eye-tracking data, significantly better than achieved through discriminant analysis (Campana, Duci, Gambini, & Scarone, 1999).

Neurologists compared a neural network to discriminant analysis to discriminate between migraine patients during attack-free periods and normal patients. Fifty-one migraine and 19 control patients were evaluated using EEG data. The ANN achieved 100% sensitivity with 15% false positives, whereas discriminant analysis achieved 73% sensitivity with 37% false positives (de Tommaso, Sciruicchio, Bellotti, Castellano, Tota, Guido, Sasanelli, & Puca, 1997).

Artificial neural network applications are gaining some popularity within the social sciences. Patterson and Cloud (2000) tested eight Bayesian ANN models to predict psychiatric re-hospitalization and found that correct prediction rates across the models ranged from 75% to 93%. Brodzinski, Crable, and Scherer (1994) compared an ANN to discriminant analysis for prediction of juvenile recidivism. The ANN correctly classified 99% of the cases, compared to 63% correct with discriminant analysis.

Research applying neural networks to CPS data is just beginning to emerge.

Marshall and English (2000) compared neural network modeling to linear and logistic

multiple regression in measuring the association between caseworkers' overall

assessments of risk and 37 separate items from the Washington Risk Assessment Matrix

(WRAM). The neural network produced superior prediction and classification results

over the regression models. The authors attributed these results as a reflection of the

superiority of the neural network for modeling nonlinear relationships between

interacting variables.

Drawbacks to ANNs

The main drawback of neural networks is in their interpretability. Regression

models can account for the incremental contributions of a particular independent variable,

while controlling for the effects of the other variables in the model. Expert systems can

be queried to account for each step involved in reaching their conclusions. But neural

networks pose what is sometimes called the "black box" problem. While they may

provide superior predictive accuracy in many cases, they are limited in their ability to

provide "explanations" regarding these relationships (Garson, 1998; Marshall & English,

2000). While some applications provide a "sensitivity analysis" (not to be confused with

sensitivity of prediction rates) that displays a rank order of the predictive factors in terms

of importance, this information is crude in comparison to the data output provided by

regression analysis.

A second problem that can arise is "overfitting" (sometimes called overtraining or

overspecification). Overfitting occurs when the network learns the noise (error) present in

the training set, in addition to the underlying relationships among the variables. Given

enough hidden nodes a neural network can be made to fit a training data set as close as

one wants. Additionally, allowing too many training *epochs* can result in overfitting.

Training "epoch" refers to a single pass of all training cases through the network. If noise

present in the training data set is memorized the network will not generalize to novel

samples (Garson, 1998).

In order to avoid overfitting during training of backpropagation networks, a

"validation training set" is used concurrently with the training data set to adjust parameter

weights. In this validation process, new cases are introduced from the validation data set

during the training stage to correct for any overspecification that may occur as patterns in

the training set are learned (Patterson & Cloud, 2000). But, even with the safeguard

provided by the validation set, overfitting may occur. The model developer must

supervise the training process and make a judgement as to when to stop training.

Typically, MLP networks are programmed to stop training when the change in error is

less than −0.01 % over five epochs (SPSS, 1997). After training is complete, a third data

set, a "test data set", may be used to test the predictive validity of the model (Garson,

1998). Lawrence (1993) recommends saving different networks during training and

testing these to determine which optimize classification in both training and testing cases.

If the final model produces highly accurate classification of cases in the training set, but

performs poorly in classifying new cases from the test data, it has likely overfitted the

training data, having learned random error patterns present in the training data. The

model in therefore useless for further application.

Also, while neural networks have been proven to outperform traditional statistical techniques in many cases, they have failed to do so for other problems. Heckert (1994) compared neural networks with discriminant analysis in prediction of firing of police personnel. The ANN did not outperform discriminant analysis, and fared worse than discriminant analysis when data were missing. Heckert concluded that it was only because the ANN accepted cases with missing data, which the discriminant analysis ignored, that the percent classified correctly was smaller for the neural network. The ANN actually classified a greater *number* of cases correctly.

Church and Curram (1996) compared neural network models with economic forecasts of the rate of growth of consumers' expenditures. The ANNs described rates as well as but no better than traditional econometrics. Dwyer (1992) compared backpropagation networks with logistic regression and discriminant analysis in the prediction of bankruptcy. The ANN outperformed discriminant analysis but was only on par with logistic regression. Pugh (1991) found that while neural networks were comparable with regression analysis in general, they did worse in predicting medical diagnoses with small data sets, where they tended to overfit the data and not generalize well. In light of such findings, Sohl and Venkatachalalm (1995) have argued that no one model provides the most accurate forecasts in all situations. Whatever the methodological procedure; the critical factor is judicious selection of the menu of independent variables. (Church & Curram, 1996).

Neural networks are not panaceas; they do not substitute for wise variable selection and accurate measurement of data. However, it can be argued persuasively that

neural network models are often superior to alternatives in terms of predictive power

(Garson, 1998). Neural networks appear to have potential for making outcome

predictions in areas where independent variables are likely to be intercorrelated and

where data are affected by moderate levels of random error and missing data. This is

likely to be the case in CPS data. However, the value of these models for risk assessment

remains speculative. Research should be directed toward testing the predictive ability of

neural networks using data that are typically available to CPS agencies.

Conclusion

Child protective service agencies have increasingly turned to formalized risk

assessment systems in response to increasing workload demands and an

acknowledgement of the need to target "high risk" cases for allocation of scarce

resources. The risk assessment systems currently in use generally fall into one of two

categories, "consensus-based" and "actuarial systems". Consensus-based systems consist

of items that are thought to be associated with risk of abuse in accordance with

etiological theories of abuse and the opinion of CPS "experts". By contrast, actuarial risk

assessment systems are composed of items obtained through examination of bivariate and

multivariate correlations among previously gathered case characteristics and subsequent

occurrences of abuse.

While risk assessment systems have proliferated widely over the past twenty

years, validation research has lagged behind. However, research examining the

performance of risk assessment systems is emerging. The performance of current risk

assessment systems can be best described as "fair", with the actuarial models tending to produce more reliable ratings, and better (yet still fairly weak) predictive validity.

Several methodological problems have hampered the development of useful predictive models. Researchers have defined the outcome variable differentially, as "subsequent report" and "subsequent substantiated incident", while others have argued that "severity" of abuse should also be considered when defining a negative outcome. These operational variations make comparisons across studies difficult. These differences in measuring outcomes also influence the "base rate", or incidence of occurrence. A more stringent criterion "only substantiated cases" reduces the base rate of interest, and thus makes prediction more difficult.

Predictor variables may interact in complex ways in relation to severity and chronicity of abuse. Several interactions have been identified, yet other influential interactions may remain unknown. Regression methods typically employed in model testing may perform poorly in the presence of such unknown interactions.

Computer-based technologies, such as expert systems and artificial neural networks have presented as potential alternatives to "traditional" statistical methods. Expert systems require the existence of human "expertise", which may be translated into a computer model for use by non-experts. This expertise does not exist in regard to making accurate predictions of recurrence of child abuse. Therefore, expert systems appear to be poorly suited for this task; however they may prove useful in other aspects of CPS work.

Conversely, artificial neural networks do not require such specialized knowledge regarding the underlying relationships among predictor variables. ANNs have been found to perform well in situations characterized by interactions, nonlinear relationships, and relatively high levels of random error. These are the conditions likely to present when examining CPS data for the purposes of developing risk assessment models. Artificial neural networks are potentially powerful tools for improving prediction risk classification in cases of physical child abuse and deserve further research to evaluate their utility in this arena.

Hypotheses

The present study proposes to develop a model from existing case file data for the prediction of the recurrence of child physical abuse, then to test the model's predictive validity on a sample of hold out data from the original data set. An artificial neural network will be tested against a logistic regression model to determine which method produces the best predictive validity. A total of 13 predictor variables and five interaction terms is selected based on findings from previous research and is discussed below. The following hypotheses will be examined:

(1) Each of the 13 predictor variables will be related to recurrence of physical abuse when controlling for all other predictor variables.

(2) When tested on hold-out data from the original set, the ANN model will be able to predict physical abuse recurrence significantly better than a logistic regression model using the same set of variables.

CHAPTER III

METHODOLOGY

Data Source

Like their civilian counterparts, military families may experience a variety of

problems, including child abuse. In response to child abuse, neglect, and spouse abuse,

the Air Force has expanded prevention and treatment services greatly over the past

twenty years (Mollerstrom, Patchner, & Milner, 1995). The Air Force Family Advocacy

Program (FAP) functions in many ways as a civilian CPS agency and seeks to:

1) Provide primary prevention services to all Air Force personnel.

2) Provide secondary prevention services to populations at risk for family

violence.

3) Support family members with special medical or educational needs.

4) Identify and treat incidents of child and spouse maltreatment.

5) Prevent child and spouse abuse.

(Air Force ServeNet, Family Advocacy Program, 2000).

Family maltreatment disrupts the military environment, drains resources, and reduces

mission readiness of military members. The goal of the FAP is to reduce the incidence of

family violence by providing services and support to families, improving family

functioning and mission readiness (Paddock, 1995).

The Family Advocacy office at each base receives allegations of abuse and

neglect from a variety of community sources including commanders, medical providers,

neighbors, and school and child care staff. Each base's Family Advocacy Program office

operates under the direction of the Family Advocacy Officer (FAO). Upon receipt of a referral, the FAO or designee reviews family medical records, contacts the family to arrange for appropriate medical appointments, notifies military law enforcement agencies, and for bases located within the U.S., coordinates with civilian CPS and law enforcement personnel (Mollerstrom, et al., 1995).

Individual and family interviews are conducted by a FAP social worker, usually the FAO. After an initial evaluation is completed, an interdisciplinary case management team is convened to determine whether the case meets criteria for substantiation, and to recommend specific treatment options to the family (Air Force ServeNet, Family Advocacy Program Standards, 2000; Mollerstrom, et al., 1995). Immediate access to the family's medical records provides valuable information (e.g. predictor variables) that is typically unavailable to civilian CPS agencies (Paddock, 1995).

The Department of Defense (DOD) (1986, 1987, 1992) requires that all military service branches collect information on reports of child maltreatment. The Air Force Central Registry located at Brooks Air Force Base in San Antonio, Texas was initiated in 1987 to collect data on all suspected cases of abuse and neglect. Data are recorded on a standard form (SF 2486, Child Abuse Incident Report), which identifies abuse type, referral source, demographics and descriptive data. If a case is deemed substantiated by the interdisciplinary case management team, more detailed information about family members is recorded (e.g. active duty member's social security number for future reference). Detailed instructions regarding completion of the form are provided to the

local Family Advocacy Programs (Air Force ServeNet, Family Advocacy Program Standards, 2000; Mollerstrom, et al., 1995).

Design

The present study is a retrospective, exploratory analysis of a large sample of substantiated and closed case reports of physical child abuse received by the U.S. Air Force's Central Registry between January 1, 1990 through June 1, 2000 (N = 5612). These case data were gathered from Family Advocacy field offices worldwide and entered into an SPSS file by researchers at the Central Registry. As will be discussed below, smaller subsets of the data will be used for selected analyses in this study.

The ten-year observation period used in this study represents the upper end of that used in previous studies of recidivism. Herrenkohl and colleagues (1979) allowed for a ten-year observation period, whereas, other studies have limited the observation period to as little as one year (Lutzker & Rice, 1987). Fluke and colleagues (1999) note that it is important for research to report the length observations periods since that they will affect observed rates of recurrence.

Two types of bias may compromise retrospective research. First, *retrospective bias* is associated with prior knowledge of the outcome of the criterion variable (Gordon, 1992). This bias is minimized in the proposed study in two ways. First, predictor variables were composed of data obtained prior to obtaining the criterion measure, independent of whether re-abuse occurred. Second, it is not the nature of the proposed machine learning method to process the validation sample any differently in regard to the

outcome, to which the machine is ignorant (Gordon, 1992). A second potential weakness

of a retrospective study is one of *sampling bias*. This problem is avoided in the present

study by including all subjects with confirmed cases of child physical abuse throughout

the Air Force over a ten-year period.

Although addressing these limitations does not equate the validity of a

retrospective design to that of a prospective design, this design has the potential to find

meaningful relationships among predictor variables and the selected criterion of repeat

physical abuse. This knowledge may prove useful in the development of future

longitudinal prospective studies. To this extent, the retrospective design is defensible for

exploratory research.

Because the data set to be used for this study is devoid of personal identifiers such

as names or social security numbers, a Form A, requesting exemption from review by the

Institutional Review Board (IRB), has been approved and is on file with the IRB. The

proposal has been approved through appropriate Air Force channels based on criteria

similar to that of the IRB.

Definitions

DOD instruction 6400.2 (1987) provides the following definitions of terms to be

used in this study. Physical child abuse is the maltreatment of a child under the age of 18

years, which results in a major or minor physical injury. A major injury includes: brain

damage, skull fracture, subdural hematoma, bone fracture, dislocations, sprain, internal

injury, poisoning, burn, scald, severe cut, laceration, bruise, welt, or any combination

thereof, which constitutes a substantial risk to the life or well-being of the victim (DOD, 1987, p, 2).

A minor physical injury includes: twisting, shaking, minor cut, bruise, welt, or any combination thereof, which do not constitute a substantial risk to the life or well being of the victim (DOD, 1987, p. 2). Note that these definitions require "demonstrable harm" in order for a case to be classified as abusive.

A *substantiated* case is a case that has been investigated and the preponderance of available information indicates that abuse has occurred. This means that the information that supports the occurrence of abuse is of greater weight of more convincing that the information that indicates that abuse did not occur (DOD, 1987; Mollerstrom, Patchner, & Milner, 1995).

Criterion Variable

The criterion measure selected for this study is recurrence of *substantiated* physical abuse involving a child who has previously suffered a substantiated case of physical abuse. These cases were identified by examining the *"masked" sponsor's social security number*, the *type of report*, *victim's date of birth*, and *incidence date* fields. The Family Advocacy Program differentiates between "repeat abuse," defined as those cases with re-abuse while the case in open, and "recidivism", defined as those cases with an case of repeated abuse after case closure. Because this distinction was not of concern to the present study, these categories were collapsed into one category to be described as either "recurrent abuse" or "recidivism" in future references.

First, examination of the masked social security number field identified duplicate social security numbers. This selected families that have had multiple reports filed with the Central Registry. Second, within this subset, those reports containing duplicate dates of birth were retained. This was done to differentiate victims of repeated abuse from families experiencing abuse of multiple victims during the initial incident. Third, the incident date field was examined to identify those cases in which a report was received after closing of a previous case. Finally, during step three, the *type of report* field was also examined to eliminate those cases that involved subsequent reports that were not indicative of a subsequent abuse incident, for example, reports of cases being transferred to other bases for continued services. This analysis of the data revealed a total of 351 cases meeting these criteria.

Because this study was concerned with those case characteristics present at the time of the initial abuse incident that predict a future occurrence of abuse, the initial case report for the re-abusing group was retained as the *index* report. The index reports for the comparison (non-recurring) group were reports of case closures of single abuse episodes.

In addition to initial substantiated case reports and repeat abuse reports, the data set contained reports on other case management activities. For instance, reports are received and recorded when open cases are transferred from one Air Force base to another or when the military sponsor separates from active service. These case reports are not necessarily examples of "non-recidivist" cases, and therefore were excluded from analysis and are not reflected in the overall N noted above.

Examination of the frequency distribution of case reports by "type of report" reveals 5,978 reports of case closings as either "resolved" (n = 5,023), indicating the case management team felt that adequate compliance with service plans had been occurred, or "unresolved"(n = 955), indicating that, for whatever reason, the service plan had not been adequately completed by the family. This distinction is not relevant to the present study. Therefore, these categories are collapsed into one category, "closed cases".

Within these 5,978 "case closure" reports are both reports of closed cases of repeat abuse, and closed cases of initial abuse. These re-abuse cases are those that were coded as "case closure" rather than either a "subsequent incident" or "case re-opened" by the reporting Family Advocacy field office. These cases were discerned from closed reports of initial abuse by examining dates of abuse incidents as described above. Of the 351 identified cases of re-abuse, 201 are coded as "case closure," and were removed from the set of cases representing non-recurring abuse.

Finally, all other case reports pertaining to those families that experienced re-abuse were purged from the data set representing non-recurring abuse. Otherwise, case characteristics representing the "recidivist" cases would be included as examples of "non-recidivist" cases. There are 437 such reports in the data set. Finally, 79 reports involving civilian families were excluded due to excessive missing data. After these exclusions, a total of 5,261 case examples of cases experiencing *no recurrence* of abuse remain.

As noted above, previous research has generally defined re-abuse as either repeat *referral* or as repeat *substantiated referral*. And as discussed above, reasonable arguments have been made for each criterion. However, the present study was limited to

substantiated recurrences of abuse, because these are the only cases on which complete information on relevant variables is collected. The result of this limitation in the data set is that the base rate of recurrence is low (6.25 % of cases with an initial substantiated case). Difficulties related to attempts to predict such rare phenomena are discussed above, and were particularly relevant to the present study. This situation provided a rigorous test of the ANN and logistic regression models to predict a low base-rate phenomenon.

Recurrence Predictors

Variable selection was driven by previous research on correlates of abuse recurrence [SEE REVIEW OF THE LITERATURE ABOVE]. Additionally, six variables not previously found to be associated with abuse recurrence, but found to correlate with related outcomes (initial abuse, severity of abuse) and available in the present data set were examined. Variable selection was constrained by the information available in the existing data set. Consistent with previous research, variables were drawn from the domains of child characteristics, caregiver/perpetrator characteristics, and situational/environmental characteristics. Data measuring seven empirically supported correlates of abuse recurrence from these domains are present in the data set. Child characteristics available include victim's age, gender, and race. Child special needs (handicapping condition) has been found to predict abuse (Burrell, Thompson, & Sexton, 1994; Weedon, Torti, & Zunder 1988). This variable is recorded by FAOs on the SF 2486. However, the Central Registry has excluded this variable from the database due to

poor reliability among field workers in recording these data (Tritt, personal

communication, December, 2000).

Relevant caregiver/perpetrator characteristics include gender, age, and substance

abuse involvement. A relevant situational/environmental variable is family income.

Considering the exploratory nature of this research, and in the interest of

furthering knowledge of predictors of physical re-abuse of children, six additional

variables (offender's marital status, offender's relationship to the child, source of referral,

severity of initial abuse, whether or not the offender received treatment, and the length of

time the case remained open) that have not been previously associated with re-abuse were

included for analysis. The *offender's marital status* was examined. Although this variable

has not been previously demonstrated to correlate with physical abuse recurrence, it is

strongly associated with initial abuse. Children of single parents are at 77 percent greater

risk of suffering initial physical abuse (NIS-3, 1996). It is also not unreasonable to

suspect that demands placed upon military single-parents (e.g. threat of sudden

deployments, geographic mobility) may be particularly stressful for these parents.

Although not reported in previous studies of recidivism, *Offender's relationship*

to the child was included in this study as an indicator of the offender's access to the child.

It is assumed that parents who have initially abused their children would have more

continued access to the victim than would extrafamilial offenders, and thus greater

opportunity to re-abuse. Therefore, the variable was dichotomized to reflect whether or

not the offender was a parent.

Source of referral was examined. Previous research has shown referral source to be strongly correlated with case substantiation decisions, with reports from anonymous or non-professional sources less likely to be substantiated. (Jason & Andereck, 1983; Zuravin, Watson, & Ehrenschraft, 1987) and with severity of injury resulting from physical abuse (Paddock, 1995; Zuravin, Orme, & Hegar, 1994). It was of interest to know if this variable would also prove to be useful for predicting case recidivism.

Severity of the initial abuse incident was considered. Previous studies have examined severity of physical abuse as an outcome variable, attempting to predict injury severity from case report characteristics (Hegar, Zuravin, & Orme, 1994; Paddock, 1995). The relationship between severity of initial abuse and future occurrences of re-abuse has not yet been reported in the literature. Whether the offender in the initial cased received *treatment* was examined in relation to future incidents of abuse. As discussed earlier, there have been mixed results in studies examining the relationship between the provision of treatment and subsequent abuse (Johnson & Clancy, 1989; Lutrell, Hull, & Wagner, 1995). Finally, the length of time that the case was open was examined. As noted above, contradictory results have been reported regarding the relationship between the amount of time a case remains open and the likelihood of recurrent abuse (Fluke, et. al., 1999; Johnson & L'Esperance, 1984).

Thus, a total of 13 predictor variables were included in the present study. These included three child characteristics (age, gender, race), five parent/perpetrator characteristics (age, gender, relationship to the victim, substance involvement, and marital status), three situational/environmental characteristics (income, source of referral,

severity of abuse), and two treatment variables (offender got treatment?, length of time case remained open).

Additionally, consistent with previous research findings, the following interaction terms were tested: child gender x child age (Kolko, 1998), race x income (Kruttschnitt, 1994), offender sex x income (Gelles 1989), and offender sex x substance involvement (Ammerman, 1999). An interaction term of victim's age x victim's age was entered to test for a curvilinear relationship between age abuse (Wolfner & Gelles, 1993).

Operational Definitions of Predictors

Perpetrator's relationship to the child was dichotomized: (0) biological parent, (1) non-parent. Child race was categorized in the original data as follows: (1) White, (2) African American, and (3) Hispanic, (4) Asian/Pacific Islander, and (5) Native American. Preliminary examination of the data showed low percentages of cases involving victims in categories 4 (4.5%) and 5 (.3%). Therefore, race was collapsed into three categories for data analysis (White, African American, Hispanic, and other minority).

Perpetrator history of substance abuse was divided into three categories: (1) alcohol and/or drug involvement, (2) no alcohol and/or drug involvement, and (3) alcohol and/or drug involvement unknown. In the original data, perpetrator's marital status was divided into four categories: (1) married (83.5%), (2) divorced (11.2%), (3) never married (4.1 %), and (4) widowed (.1%). Because of the small percentages representing categories 3 and 4, categories 2, 3, and 4 was collapsed into one category, "not currently married", thus dichotomizing this variable. The treatment variable is dichotomous, with

"perpetrator got treatment" coded as "1", and "perpetrator did not get treatment" coded as "0". The *time case remained open* variable is continuous, measured in months.

Severity of maltreatment is recorded by Family Advocacy Officers as a three-tier ordinal-level variable: (1) mild, (2) moderate, and (3) severe. All case managers receive annual training on the use of severity ratings. Mollerstrom (1989) found 95 % agreement among a sample of Family Advocacy Officers using this scale. However, this study reported observed agreement, and did not correct for chance agreement. Therefore this estimate of interrater reliability is likely inflated. Therefore, the reliability of this variable is unknown. Due to the small number of categories, severity of maltreatment will be treated as a categorical variable.

Referral source was originally divided into eighteen categories. To eliminate categories with very small frequencies and to be of practical use, these data were collapsed into six categories: (1) law enforcement, (2) medical/dental provider, (3) school/child care personnel, (4) civilian CPS, (5) other professional, (6) other non-professional. Family income is reported as a ratio variable, using the number of dollars per year. Income was estimated on the basis of the sponsor's (active duty service member) rank and military benefits (e.g. housing and separate rations allowance), and thus does take into account part-time jobs or spousal employment. Since beginning salaries for each pay grade were used, income is likely somewhat underestimated across pay categories. Perpetrator's age and child's age were treated as quantitative variables. Finally, perpetrator's gender and child's gender are dichotomous variables.

Missing Data

A data check and missing data analysis were done to determine if any data entry errors had occurred and to examine percentages and patterns of missing data. The test revealed that twelve of the 13 predictor variables had less than one percent missing data. One variable, offender's age, had 7.1 percent missing data. A data field is usually considered useful if 70 % or more of the records contain values (SPSS, 1997). The missing value analysis revealed no patterns in missing values across cases.

Considering the relatively small amount missing data, and the small frequency of cases in the re-abuse category of the criterion variable, data replacement was used in lieu elimination of cases with missing data points for the offender's age and victim's age variables, thus retaining the maximum number of cases for network training and testing. There were no significant differences between estimated age values obtained through regression analysis and simple group mean computation. The regression-derived estimate was chosen to replace missing data points for the offender's age variable because it uses a more available information to predict the expected value than does mean substitution.

Additionally, a small number of obvious coding errors were identified in the *offender's age* (16) and *victim's age* (37) fields. An entry in one these fields was considered erroneously coded if it contained a negative number or when the recorded age was outside the range defined by Air Force Instructions, for instance a victim above the age of eighteen years. These values were treated as missing and were replaced with imputed values.

Parameter estimates were obtained using the full data set (N = 5612) with and without imputed values for the two age variables. No significant differences in regression slopes or odds ratios were observed. Regression results reported were obtained from the data set with imputed values for the age variables.

Sampling Considerations

The data was divided into two subsets: The *training data* subset was used during the model building phase of the analysis to compute model weights, and a *test data* subset that is used to test model generalization after training is completed. There is no clear consensus within the neural network literature in regard to the optimal division of data into training and test data sets. However, in discussions of cross-validation of prediction and classification models using more traditional statistical analyses (e.g. regression) several authors have recommended dividing data into disproportionate sub-samples, with significantly more data assigned for model development, thereby maximizing the amount of data available for parameter estimation (Cooil, Winer, & Rados, 1987; Picard & Cook, 1984; Thompson, 1994). One convention for ANN training is to retain a minimum of $10(M+N)$ cases for network training, with N indicating the number of independent variables, and M indicating levels of the dependent variable. In the present study, after dummy coding of variables described below, there are 30 input variables and 1 dependent variable. Thus, $10(30+1) = 310$. However, when plenty of data is available, it is better to use more for training (SPSS, 1997).

In a study to predict psychiatric re-hospitalization, Patterson and Cloud (2000) trained eight Bayesian ANN models, varying the allocation of data set aside for model testing (30%, 40%, 50%, and 60%). The model that set aside 30 percent of the data for testing, thus training on 70 percent of the data, produced superior accuracy in the prediction of the dichotomous outcome of re-hospitalization.

Although the present study employs a relatively large N (5,261), the number of cases available to train the ANN on input patterns relevant to "recidivist" cases is small (n=351, or 6.25% of total N). Training an ANN with such a disproportionate representation of cases across the target variable (e.g. re-abuse versus no re-abuse) can cause the network to learn to predict all case as belonging to the larger group instead of learning features that discriminate between the groups (SPSS, 1997). To avoid this SPSS recommends equalizing the ratio of "responders" to "non responders" during the training phase (p.11). In the current study, this entailed randomly selecting 70% (246) of recidivist cases, then randomly selecting an equal number of non-recidivist cases, for a total N of 492. Thus, what at first appeared to be a very large data sample was shrunk to a sample just adequate for the proposed analysis. This again illustrates the difficulties related to the low base-rate recidivist cases. To maintain consistency of procedures for the ANN and logistic regression models, each method was the smaller set containing an equal ratio of re-abuse and non-re-abuse (N = 492) for parameter estimation. Classification results reported are based on model performance on the *test data set* (N = 1683), which represents the actual recurrence rate of 6.25%.

One may increase the amount of data available for training by duplicating the "positive" cases, allowing for use of a larger number of "negative" examples in the training process, and consequently improving network performance (SPSS, 1997). Since the number of cases available for training in the current study exceeded the recommended minimum, this duplication method was not used.

Data Analysis

The data analysis consisted of five parts. First, frequency distributions and descriptive statistics were computed to check for data entry errors or other anomalies in the data. Second, bivariate correlations between potential predictors and abuse recurrence were calculated using cross-tab tables for categorical independent variables and Analysis of Variance for continuous independent variables. Logistic regression was used to test hypothesis 1, that each of the 13 predictor variables is related to the outcome of re-abuse when controlling for all of the other predictor variables in the model. The full data set (N = 5612) was used for bivariate and multivariate tests of the individual predictor variables.

These analyses provided information regarding characteristics that distinguish between "recidivist" and "non-recidivist" cases. Logistic regression is a kind of regression analysis often used when the dependent variable is dichotomous and scored 0 or 1. It is usually used for predicting whether or not something will happen, and can be used with categorical or continuous independent variables. Logistic regression has become increasingly used as an alternative to discriminant analysis because it requires fewer assumptions (Vogt, 1999). ANNs have been tested against of logistic regression

models in several medical outcome studies (see review of the literature). All bivariate and logistic regression analyses were conducted using SPSS 10.0 statistical software (1999).

Third, the BLR model was re-estimated using the "estimation data set" (N = 492) to produce parameter estimates for the prediction model. Fourth, an ANN model was also built, using the estimation data, for the purpose of predicting recurrence of physical abuse The ANN model was "trained" on this subset of data, identifying relationships among the selected predictor variables and the target variable of recurrence of abuse. Fifth, to test the predictive validity of the ANN model against that of the logistic regression model (Hypothesis 2), each model was tested for accuracy using the cross validation or "test" data set that was not included in the model development and training process.

Receiver Operating Characteristic (ROC) Curve Analysis was used to examine the predictive accuracy of the two methods. As discussed above, ROC curve analysis allows for examination of predictive accuracy across a spectrum of diagnostic cut points. ROC curves are able to quantify the accuracy of classification systems without regard to the probability distributions of training and test set pattern vectors (Zaknich, 1998). One can visually examine the relative cost, in increasing *false positive* predictions, resulting from relaxing classification criteria in favor of detecting *true positive* cases. The area under the ROC curve varies between 0.5, representing no discrimination, to 1.0 representing prefect discrimination (Zaknich, 1998). In addition to visual inspection, the area under the ROC curve can be tested in regard to being significantly different from chance. Also, the ROC areas produced by two instruments or methods can be tested to

determine whether they are statistically different from one another. A Wilcoxon two-sample Z statistic can be used for this purpose (Hanley & McNeil, 1982, 1983).

Neural Network Training

All ANN applications were conducted on a 400 MHz personal computer. A commercially available neural network software package (Neural Connection 2.0, SPSS, Inc.) was used to run all ANN analyses. A Bayesian network, which uses a backpropagation, feedforward architecture and training algorithm similar to that of the MLP described above was used. The Bayesian model, which is architecturally similar to the MLP, was chosen because it is useful for prediction and classification problems (SPSS, 1997) like the one presented in the present study and because it presents advantages over the standard MLP.

The most significant advantage of the Bayesian network is that the algorithm used does not require a *validation* data set to be used during the training process. Unlike the *test data set*, which is used to test generalizability of the model after model development and training, the *validation data set* is used in an iterative fashion during training to prevent overtraining of a standard MLP neural network (Garson, 1998). The Bayesian network automatically prevents overtraining by adding an additional term to the learning algorithm that corrects for the effect of error/noise in the data (Patterson & Cloud, 2000), thus preserving more data for model training and testing. The additional term used in Bayesian models is derived from the application of Bayesian statistics, which examine how prior and posterior knowledge of events affect probability distributions and

subsequent prediction error (Patterson & Cloud, 2000). Bayesian models are especially

useful when data are sparse (Garson, 1998), as is the case in the present study with regard

to recidivist examples.

Architecture

Input Layer

The input layer of the ANN is comprised of one "node" for each continuously

valued predictor variable such as income, taking on the value of the measure. Rank-

ordered categorical predictor variables were also represented with one node per predictor

variable, taking on the number value of the interval category. Discrete categorical

variables were represented with a group of nodes, one node per level of the variable, with

each node taking on a value of "1" to indicate category membership, and "0" indicating

non-membership. Dichotomous predictor variables were coded "1" or "0" to indicate

presence or absence of the variable (e.g. Offender's marital status). After re-coding there

were a total of 26 input nodes.

Hidden Layer

Neural Connection allows the user to select one or two hidden layers of neurons,

however it has been demonstrated that one hidden layer is sufficient for most

applications, and additional hidden layers may increase the risk of overtraining (Garson,

1998; SPSS, 1997). The user can also set the number of nodes in the hidden layer.

Increasing the number of nodes will likely improve network performance in the training

data, but may also lead to poor generalizability to the test data (SPSS, 1997). There is no

well-established rule for determining the optimal number of hidden nodes. If too few

nodes in the hidden layer may cause the model not to train well. Whereas too many nodes

may result in overtraining (Garson, 1998). A very general rule is to take the number of

input neurons and number of output nodes, average them, and use this number as the

number of hidden nodes (Garson, 1998). This convention was used as a starting point for

the current study, establishing one hidden layer of 13 nodes. However, ANNs are

developed in an experimental manner to increase predictive ability and reduce error

(Patterson & Cloud, 2000). Therefore, the number of nodes in the hidden layer may be

increased (grown) or decreased (pruned) during training in order to improve performance.

<u>Output Layer</u>

The output (or target) layer in the present study is dichotomous. Cases in which

re-abuse has occurred were coded "1" and those in which no re-abuse has occurred were

coded "0". The initial ANN model consisted of 26 input nodes, one layer of 13 hidden

nodes, and 1 output node: See figure 1, below. The best neural network produced in the

present study had 13 input nodes after variable elimination discussed below, and 4 nodes

in the hidden layer, thus a final architecture of 13-4-1. [See Figure 1 – Artificial Neural

Network Model below].

In summary, 13 independent variables available in the current data set were

examined in relation to outcome of repeated physical abuse in families with a previous

case of substantiated physical abuse. From this set of variables a prediction model was

developed and tested against randomly selected set of "hold out" cases. The predictive

accuracy of an artificial neural network was compared to the more "traditional" analytic

method, logistic regression.

Figure 1 – Artificial Neural Network Model

CHAPTER IV

RESULTS

Descriptive Statistics and Bivariate Analyses

Data from a total of 5,612 cases were analyzed, in which 5,261 (93.7%)

experienced only single incidents of abuse, and 351 (6.3%) in which a subsequent

incident of physical abuse was documented. See Table 3 – Cross-tabulation Results, and

Table 4 – AVOVA Results for descriptive data for each group. Bivariate correlations are

reported for the purpose of providing a general overview of differences observed across

the two groups. As will be seen, some of the observed bivariate relationships became

non-significant in regard to predicting re-abuse in multivariate analysis. Also, examining

13 separate bivariate relationships increased the overall risk of a type I error to .49,

meaning that the chance of finding at least one significant correlation due to chance is

49%.

Bivariate analyses revealed significant group differences ($p \leq .05$) for six of the

13 predictor variables. *Referral source* was associated with re-abuse, with those children

experiencing a subsequent abuse incident experiencing a change in risk by a factor of .59

(41% *less likely* to have been initially referred by a medical provider, as compared to the

other referral sources) (O.R. = .59, X^2 = 10.84, p < .01). There were no significant

findings in regard to the other five referral sources.

Victim's race was associated with recurrence of physical abuse, with the *other*

racial group experiencing a change in risk by a factor of .44. Those children who

experienced subsequent physical abuse were 56% less likely to be from the *Other Race*

Table 3- Crosstabulation Results (N=5612)

Predictors	Reabuse % n=351	No Reabuse % n=5261	Odds Ratio	X^2	p
Severity					
High	6.8	5.1	1.36	1.98	.16
Medium	25.4	29.2	.825	2.34	.13
Low	67.8	65.7	1.1	.64	.42
Referral Source					
Law Enforcement	13.4	13.3	1.01	.003	.96
Medical	13.4	20.7	.59	10.84	<.01*
CPS	14.8	17.3	.836	1.43	.23
School	24.5	20.3	1.27	3.49	.06
Other Professional	5.6	20.3	1.27	3.49	.06
Other Non Professional	5.6	5.5	1.39	2.38	.12
Victim Female	48.4	46.5	1.08	.51	48
Offender Got Treatment	65.8	62.3	1.17	1.75	.19
Victim Race					
White	69.4	68.4	1.05	.15	.67
Black	24.9	20.8	1.26	3.26	.07
Hispanic	3.4	5.7	.58	3.32	.07
Other	2.3	5.0	.44	5.37	.02*

Table 3 - Continued

Predictors	Reabuse %	No Reabuse %	Odds Ratio	X^2	p
Offender Sex					
Female	35.3	35.6	.99	.01	.93
Offender Marital Status Single	17.1	15.7	1.1	.45	.50
Offender Relationship Non-Parent	4.8	8.5	.55	5.7	.02*
Substance Involvement					
Yes	3.4	4.8	.71	1.31	.25
No	81.2	76.8	1.30	3.53	.06
Unknown	15.4	18.4	.81	2.0	.16

*$p \leq .05$

Table 4- ANOVA Results (N=5612)

Predictors	Reabuse		No Reabuse				
	__M__	__SD__	__M__	__SD__	F	df	p
Offender's Age	30.4	5.7	31.3	6.8	5.3	1	.02*
Victim's Age	8.0	4.8	8.4	5.0	1.9	1	.17
Time Case Open (months) (n=5603)	8.6	7.3	5.8	5.0	96.6	1	<.01*
Family income	26,477	4,620	27,480	5,261	10.8	1	<.01*

*p ≤ .05

group (primarily Asian) as compared to White, African-American, and Hispanic children ($O.R.=.$ 44, $\underline{X}^2 = 5.37$, $\underline{p} = .02$). There were no significant bivariate differences among White, African-American, and Hispanic children in regard to recurrent abuse.

Having been initially abuse by a non-parent perpetrator was associated in a change in risk by a factor of .55. Victims experiencing recurrent abuse were 45% less likely to have suffered the initial abuse incident at the hands of a *non-parent perpetrator.* Stated differently, those children who were eventually re-abused were 55% more likely to have been initially abuse by a parent ($\underline{O.R.} = .55$, $\underline{X}^2 = 5.7$, $\underline{p} = .02$).

Offender's age at the time of the initial abuse incident was related to subsequent abuse. Children who were subsequently re-abused were initially abused by younger perpetrators ($\underline{M} = 30.4$, $\underline{SD} = 5.7$) than were those who suffered no recurrence of abuse ($\underline{M} = 31.3$, $\underline{SD} = 6.8$, $\underline{p} = .02$).

The child's family income was related to subsequent abuse. Those suffering a recurrence of abuse were from slightly lower income homes ($\underline{M} = \$26,477$, $\underline{SD} = 4,620$) compared to those suffering no abuse recurrence ($\underline{M} = \$27,480$, $\underline{SD} = 5,261$, $\underline{p} < .01$). As noted above, the income variable only accounted for the income of the military member. Finally, the length of time that the initial case remained open was related to abuse recurrence. Cases in which re-abuse occurred had been kept open longer in response to the initial incident ($\underline{M} = 8.6$ months, $\underline{SD} = 7.3$) than those that did not experience subsequent abuse ($\underline{M} = 5.8$ months, $\underline{SD} = 5.0$, $\underline{p} < .01$).

All other predictor variables: severity of initial injury, victim's gender, victim's age, offender's gender, offender's marital status, substance involvement, and whether or not the offender got treatment were not found to be associated with recurrence of physical abuse at a bivariate level.

Regression Results

Binary logistic regression (BLR) was conducted on two samples. First, parameter estimates were computed on the full data set (N= 5612), then estimates were obtained using a sample made up of 70% of the "re-abuse" group and an equal number of randomly selected cases from the "no re-abuse" group (N = 492). These data will be referred to as the "estimation data set". The first sample method was chosen in order to maximize the number of cases available for parameter estimation. The second, smaller sample was chosen in order to provide a consistent comparison with the ANN model, which as discussed above, should be trained on equal proportions of each level of the target variable (SPSS, 1997). Examination of the results reveals similar results for the two samples (See Tables – 5 and 6 Logistic Regression Results for Full Data Set and Tables 7 and 8 – Logistic Regression Results for Estimation Sample).

In addition to the 13 predictor variables discussed above, five interaction terms were created and entered into the BLR model. Selected interactions were based on previous research findings discussed in the review of the literature section, and included: child's age x child's age (to test for a curvilinear relationship), child's sex x child's age, child's race x family income, offender's sex x income, and offender's sex x substance

TABLE 5 - Logistic Regression Results

MAIN EFFECTS – Full Data Set (N =5612)

	B	S.E.	Wald	df	Sig.	Exp(B)
victim age	.012	.016	.565	1	.452	1.012
offender age	-.014	.013	1.231	1	.267	.986
time case open	.075	.009	76.276	1	.000	1.078
income	-.00003	.000	4.923	1	.027	.99997
severity			2.935	2	.230	
medium	-.179	.108	2.716	1	.099	.836
high	.163	.160	1.032	1	.310	1.177
victim race			10.354	3	.016	
Af. American	.460	.146	9.903	1	.002	1.585
Hispanic	-.272	.243	1.250	1	.264	.762
other	-.484	.284	2.906	1	.088	.616
offender female	-.046	.121	.143	1	.705	.955
victim female	.143	.116	1.535	1	.215	1.154
offender not married	.194	.151	1.642	1	.200	1.213
substance involvement			3.277	2	.194	
no	.214	.120	3.175	1	.075	1.239
unknown	.015	.145	.011	1	.917	1.015
offender non-parent	-.377	.282	1.788	1	.181	.686
referral source			14.025	5	.015	
medical	-.424	.142	8.950	1	.003	.654
school	.158	.116	1.873	1	.171	1.171
cps	-.187	.136	1.867	1	.172	.830
other prof.	.262	.184	2.016	1	.156	1.299
other non-prof.	.141	.110	1.628	1	.202	1.151
offender got treatment	.046	.120	.146	1	.703	1.047
Constant	-2.485	.453	30.117	1	.000	.083

TABLE 6 - Logistic Regression Results with Interaction Terms

Full Data Set (N=5612)

	B	S.E.	Wald	df	Sig.	Exp(B)
victim age	.078	.044	3.143	1	.076	1.081
offender age	-.012	.013	.835	1	.361	.988
time case open	.075	.009	77.679	1	.000	1.078
income	.000	.000	11.532	1	.001	1.000
severity			3.420	2	.181	
medium	-.199	.109	3.304	1	.069	.820
high	.202	.163	1.529	1	.216	1.224
victim race			9.369	3	.025	
Af. American	.735	.283	6.743	1	.009	2.086
Hispanic	-.483	.306	2.495	1	.114	.617
other	-1.326	.773	2.941	1	.086	.266
offender female	1.777	.806	4.862	1	.027	5.912
victim female	-.039	.227	.029	1	.865	.962
offender not married	-.189	.152	1.544	1	.214	.828
substance involvement			3.016	2	.221	
no	.204	.121	2.847	1	.092	1.226
unknown	.115	.171	.449	1	.503	1.121
offender non-parent	.353	.283	1.549	1	.213	1.423
referral source			13.728	5	.017	
medical	-.417	.142	8.562	1	.003	.659
school	.131	.117	1.254	1	.263	1.140
cps	-.198	.137	2.082	1	.149	.820
other prof.	.269	.185	2.113	1	.146	1.309
other non-prof.	.149	.111	1.821	1	.177	1.161

TABLE 6 (Continued)

Offender got treatment	-.040	.121	.108	1	.743	.961
race x income	.000	.000	1.348	1	.246	1.000
victim age x victim age	-.004	.002	2.887	1	.089	.996
victim sex x victim age	.013	.025	.280	1	.596	1.013
Offender sex x income	.000	.000	8.474	1	.004	1.000
Offender sex x substance	-.302	.300	1.015	1	.314	.739
Constant	-4.264	.857	24.771	1	.000	.014

a Variable(s) entered on step 1: RACINC, VAGECURV, VSEXAGE, OSEXINCO, SEXSUBST.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.871	8	.662

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	**Step**	15.870	5	.007
	Block	15.870	5	.007
	Model	142.227	26	.000

Table 7 – Logistic Regression Results

Main Effects – Estimation Set (N = 485)

	B	S.E.	Wald	df	Sig.	Exp(B)
offender age	-.003	.026	.016	1	.900	.997
victim age	-.024	.027	.742	1	.389	.977
time case open	.107	.020	27.885	1	.000	1.112
income	-.00006	.000	4.614	1	.032	.99994
severity			.171	2	.918	
medium	-.064	.207	.095	1	.758	.938
high	.032	.329	.009	1	.923	1.032
victim race			7.661	3	.054	
Af. American	.667	.246	7.344	1	.007	1.948
Hispanic	-.061	.366	.028	1	.867	.941
Other	-.805	.468	2.951	1	.086	.447
offender female	-.254	.210	1.469	1	.226	.776
victim female	.548	.210	6.798	1	.009	1.730
offender not married	.052	.276	.036	1	.850	1.053
substance involvement			1.906	2	.385	
no	.164	.197	.697	1	.404	1.178
unknown	-.222	.239	.866	1	.352	.801
offender non-parent	.067	.433	.024	1	.876	1.070
referral source			9.591	5	.088	
medical	-.479	.250	3.669	1	.055	.619
school	-.284	.199	2.024	1	.155	.753
CPS	.130	.251	.266	1	.606	1.138
other prof.	.750	.383	3.846	1	.050	2.118
other non-prof.	.172	.198	.755	1	.385	1.188
offender got treatment	-.172	.210	.669	1	.413	.842
Constant	.788	.799	.973	1	.324	2.200

Table 8 - Logistic Regression Results with Interaction Terms

Estimation Set (N=485)

	B	S.E.	Wald	df	Sig.	Exp(B)
Offender age	-.001	.026	.003	1	.960	.999
victim age	.068	.089	.591	1	.442	1.071
time case open	.095	.020	23.366	1	.000	1.099
Income	.000	.000	9.890	1	.002	.99989
Severity			.148	2	.929	
medium	-.073	.211	.120	1	.729	.930
high	.130	.341	.145	1	.703	1.138
victim race			9.190	3	.027	
Af. American	1.191	.507	5.525	1	.019	3.291
Hispanic	-.451	.504	.801	1	.371	.637
Other	-2.289	1.359	2.839	1	.092	.101
Offender female	-1.423	.718	3.927	1	.048	.241
victim female	.278	.210	1.760	1	.185	1.320
Offender not married	.055	.141	.154	1	.694	1.057
Substance involvement			.629	2	.730	
no	.156	.198	.615	1	.433	1.168
unknown	-.006	.297	.000	1	.985	.994
Offender non-parent	.055	.224	.061	1	.805	1.057
referral source			8.147	5	.148	
medical	-.490	.251	3.790	1	.052	.613
school	-.251	.202	1.545	1	.214	.778
CPS	.128	.257	.248	1	.618	1.137
other prof.	.645	.381	2.868	1	.090	1.905
other non-prof	.171	.200	.728	1	.394	1.186
Offender got treatment	-.062	.107	.338	1	.561	.940

Table 8 (continued)

victim age x victim age	-.005	.005	1.095	1	.295	.995
victim sex x victim age	.000	.043	.000	1	.991	1.000
offender sex x substance involvment	-.493	.477	1.066	1	.302	.611
offender sex x income	.000	.000	5.800	1	.016	1.00012
victim race x income	.000	.000	1.159	1	.282	1.000
Constant	-.457	1.231	.138	1	.710	.633
a Variable(s) entered on step 1: VAGEVAGE, VSEXAGE, SEXSUBS, SEXINC, RACEINC.						

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	11.411	8	.179

involvement. (See Table 5 – Logistic Regression Results – Main Effects and Table 6 –

Logistic Regression Results – Interaction Terms).

Results reported for individual variables are derived from full data set (N=5612).

The results of the binary logistic regression reveal that the overall model was statistically

significant (overall $X^2(26) = 142.227$, p <.01; Hosmer-Lameshow (8) $X^2 = 5.87$, p

=.662) (Note: non-significance is indicative of good model fit with the Hosmer-

Lameshow statistic). When all predictors and selected interactions were entered into the

regression as control variables five were statistically significant at the .05 level. These

were: length of time the initial case was kept open, income, victim's race, source of the

initial referral, and one of the interactions; offender's sex x income.

The *length of time* the initial case was monitored was positively correlated

associated with risk of future re-abuse. A one-month increase in the time the case was

monitored was associated with an increase in risk of re-abuse by a factor of 1.08 (8 %).

The victim's household *income* was negatively correlated with risk of re-abuse. A one

thousand-dollar increase in family income was associated with a change in risk for re-

abuse by a factor of .99997 (4%). Although associated with an interaction with offender's

sex, this main effect is reported due to the fact that this interaction was very weak. In fact,

the differential effect of income for males and females could not be represented

graphically without extrapolating beyond available data. To further explore the

relationship between offender sex and income, separate analyses were conducted for

males and females to in regard to the main effect of income on re-abuse.

In these analyses, the main effect for income was significantly associated with re-abuse when the perpetrator was male (n=3617, O.R.=.93, C.I.=.89 - .97, p < .01). When only those cases involving a female offender were examined, the main effect for income on re-abuse was non-significant (n=1995, O.R.=1.0, C.I.=.97 – 1.1, p=.6).

The *victim's race* was associated with risk of re-abuse. African-American race was associated with an increase in risk for re-abuse by a factor of 1.59. African-American children were 59% more likely to experience re-abuse, as compared to all racial groups combined. When data was re-analyzed using *White* children as the reference group, as opposed to the overall mean for all groups combined, only the *other racial group* (primarily Asian) was significantly different from the White children. In this analysis, children from the other racial group were less likely to suffer re-abuse by a factor of .46 (54% less likely) (O.R. = .46, C.I. = .222 - .946, p = .04).

Of the six *referral sources*, one was significantly correlated with re-abuse. Referral from a medical provider was associated with a decreased risk for re-abuse by a factor of .65. Those children who were initially referred by a medical provider were 35% *less* likely to suffer an abuse recurrence, as compared to all groups combined.

As noted above, one of the interaction terms were significant at the .05 level. The offender's sex x family income was significant, with an increase in income more strongly associated with a decrease in abuse risk when the perpetrator if the initial abuse was male. It should be noted that this interaction, while statistically significant, were very weak and their influence was small in terms improving the predictive validity of the overall model.

When parameters were re-estimated on the much smaller "estimation sample," some minor differences were noted. Five predictors were statistically significant at the .05 level: Time the initial case remained open, referral source, victim race, offender sex, and household income. *Offender's sex*, which was non-significant in when calculated on the estimation sample (p = .08) became significant at the .05 level when calculated on the smaller sample. Other significant predictors were consistent across the two samples.

Because estimates derived from the larger data set are assumed to be more precise than are those from the smaller estimation sample, results regarding individual predictors from the estimation sample were not interpreted. The purpose of the estimation sample was to provide a balanced comparison between the BLR model and the ANN, which requires equal representation of categories of the dependent variable during the training phase.

Results of the multivariate analysis reveal that fewer than half of the selected predictors are correlated with the outcome of recurrent abuse within this sample. Therefore, **Hypothesis 1**, that each of the 13 predictor variables would be correlated with recurrence of physical abuse when controlling for all other predictor variables was not confirmed.

Variable Elimination

In the interest of parsimony and maximization of statistical power, non-significant variables were not included in the prediction models. ROC Curves were calculated using the all 13 main effects variables and five interaction terms and then with a smaller model

retaining five main effects variables and one interaction. Inclusion of the twelve non-significant variables resulted in negligible improvement in predictive validity.

When calculated on the "estimation sample" (N = 492), the respective ROC areas were not significantly different (18 variable model ROC area = .728; 6 variable model ROC area = .712, $\underline{Z}$ = -.681, $\underline{p}$ = .496). When calculated on the full data set (N = 5612) the difference in ROC areas was statistically significant (18 variable ROC area = .681; 6 variable model ROC area = .669, $\underline{Z}$ = -4.854, $\underline{p}$ < .001). Although statistically significant, this difference translated into very little practical significance in regard to prediction accuracy within the estimation sample. Therefore, the more parsimonious model was chosen for model development and testing. Retained variables were: Time case kept open, income, victim's race, referral source, offender's sex, and the offender sex x income interaction. As discussed above, offender's sex was non-significant in terms of main effects. It was retained in the final model due to its role in the significant interaction term.

BLR Model versus ANN Model

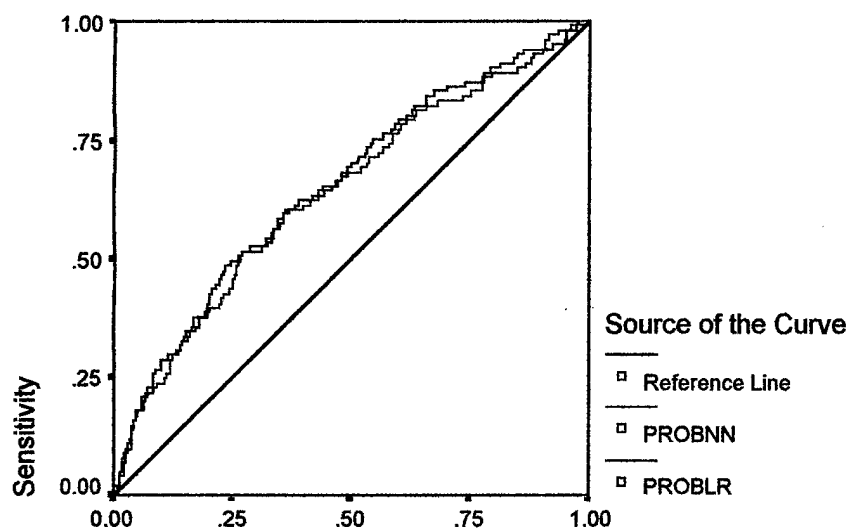<u>Logistic Regression Prediction Model</u>

When applied to the hold-out data (N = 1,679), the BLR model predicted abuse recurrence significantly better than chance (ROC area = .65, C.I. = .592 - .708, $\underline{p}$. < .001) (See Figure 2 – ROC Curves). At the .5 threshold, the BLR model correctly classified 61% of all cases; 63 of 101 re-abuse cases (62% sensitivity), and 960 of 1,578

Case Processing Summary

DEP_VAR	Valid N (listwise)
Positive	101
Negative	1578

ROC Curve

1 - Specificity

Diagonal segments are produced by ties.

Area under the Curve

Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
PROBLR	.650	.030	.000	.592	.708
PROBNN	.644	.029	.000	.587	.702

Figure 2 - ROC Curves

non-recurring cases (61% specificity) (Kappa = .063, p = <.01). At the .4 threshold the

model correctly classified 82% of Re-abuse cases but at the cost of incorrectly classifying

65% of ` of non-recurring cases as recurrent. At the .6 threshold only 38% of recurrent

cases were identified, while correctly classifying 82% of non-recurring cases.

Neural Network Model

When applied to the holdout data (N = 1,679), the ANN model also predicted

abuse recurrence significantly better than chance (ROC area = .644, C.I. = .587 - .702, p

< .01). At the .5 threshold, the ANN model correctly classified 63% of all cases; 61 of

101 re-abuse cases (60% sensitivity), and 1,005 of 1,578 non-recurring cases (64%

specificity) (Kappa = .069, p = <.01). At the .4 threshold the model correctly classified

78% of re-abuse cases, but at the cost of incorrectly classifying 61% of non-recurring

cases as recurrent. At the .6 threshold only 37 of recurrent cases were identified, while

correctly classifying 83% of non-recurring cases.

Contrary to research hypothesis 2, the ANN model did not predict abuse

recurrence better than the BLR model across the range of diagnostic thresholds. Although

generally comparable to the ANN, the BLR model produced a larger area under the ROC

curve. Although small in terms of practical utility, this difference was statistically

significant (ROC Area BLR = .650, ROC Area ANN = .644, Z = -5.782, p < .01). (See

Table 9 – Wilcoxon Matched-Pair Signed Ranks Test).

To examine the possibility that the variable elimination procedure may have

eliminated variables that, while non-significant in the regression analysis, could have

played roles in higher order interactions, the ANN model was re-trained with all variables

Table 9 - Wilcoxon Matched-Pair Signed Rank Test

Ranks

		N	Mean Rank	Sum of Ranks
PROBNN - PROBLR	**Negative Ranks**	951	862.33	820077.00
	Positive Ranks	728	810.83	590283.00
	Ties	0		
	Total	1679		

Test Statistics

	PROBNN - PROBLR
Z	-5.783
Asymp. Sig. (2-tailed)	.000

included. Inclusion of all 13 predictor variables resulted in *poorer* ANN performance. The area under the ROC Curve became non-significant (area = .56, $\underline{p}$ = .06). At a threshold of .5, prediction accuracy dropped to 54% overall, 53% sensitivity and 55% specificity (Kappa = .02, $\underline{p}$ = .11).

Finally, numerous variations of the number of nodes in hidden layer of the Bayesian network failed to improve the results. The best performing ANN model was that with 4 nodes in the hidden layer, produced by the automatic node generation option of the SPSS Neural Connection program.

In summary, models produced by BLR and Bayesian ANN predicted physical child abuse recurrences better than chance. The models were similar in terms of predictive accuracy across various diagnostic thresholds. The BLR model produced slightly better predictions than did the ANN. **Hypothesis 2**, that the Bayesian ANN model would demonstrate better predictive validity than the BLR model was not confirmed.

CHAPTER V

DISCUSSION AND IMPLICATIONS

Research Hypotheses and Previous Research

Child Protective Services have increasingly begun to turn to formal risk assessment systems to augment intuitive judgement, which may be unreliable in the prediction of subsequent abuse and neglect. Predicting abuse recurrence appears to be a difficult task and current models have received limited empirical support. However, caseworkers are making critical decisions regarding abuse cases every day and the current dearth of research on these risk assessment models should serve as a call for more research, not as a justification to abandon the effort.

The present study sought to systematically develop and train an artificial neural network model from case data to predict recurrences of child physical abuse, to test this model against a sample of hold-out cases; and then to compare the predictive accuracy of the ANN model to that produced by a more traditional statistical method, binary logistic regression. 13 variables were identified within the data as potential predictors of abuse recurrence.

Hypothesis 1, that each of the selected predictor variables would be correlated with the outcome variable when controlling for all other variables in the model was not confirmed. Of the 13 predictors, only four were correlated with re-abuse at the .05 level. These were: victim's race, household income, source of the initial referral, and the length of time the case remained open after the initial referral. When parameters were estimated

on the smaller "estimation sample" the age of the offender was also significantly related to the dependent variable. Because this relationship became non-significant when the larger estimation set was used, it is most likely a result of sampling error in the smaller sample.

In addition to the individual predictor variables, one interaction term was correlated with re-abuse at the .05 level. The offender's gender interacted with household income, with higher income being slightly more protective when the offender is male. Each of the predictor variables, starting with those that showed a statistically significant correlation with re-abuse, are discussed below in relation to findings from previous studies.

Child (victim) Race

Initial analysis revealed that African-American children were 1.59 times more likely to suffer re-abuse, as compared to all racial groups taken as a whole. This finding is inconsistent with a Children's Research Center study, which found no significant differences in re-abuse rates among White, African-American, and Hispanic children (Baird, et al, 1999).

When data were re-analyzed, using White as the reference group, only the *other racial group* (primarily Asian) was found to differ significantly from White children. Children from the *other group* were 56% less likely to suffer re-abuse than were White children. This analysis, along with the bivariate analysis of race is consistent with a previous study that reported that Asian families demonstrated a longer survival period until recidivism compared to the three other groups (Fluke, et al. 1999).

As noted above, previous research has shown cases involving African-American families are more likely to be found substantiated by CPS agencies despite national research indicating similar rates of maltreatment for African-Americans and Whites (Baird, 1999). Because the present study used *only* substantiated initial and repeat incidents, findings from the initial data analysis regarding this variable could reflect institutional bias in the way case status determinations are made for different racial groups rather than actual increased risk experienced by African-American children. African-American children may have been over-represented in the data due to differences in the ways substantiation decisions were made for different racial groups.

Family Income

Children who suffered re-abuse came from families with slightly lower income (M = $26,477) than those suffering no re-abuse (M = $27,480). As noted above the present study was only able to account for the military member's income. Therefore, these estimates are likely deflated. Also, there was no available information regarding differences between one-income and two-income families in relation to re-abuse. In the regression model, a $1,000 increase in income was associated with a 4% reduction is risk for re-abuse.

Socioeconomic status has been consistently found to correlate with abuse. Sedlack and Broadhurst (1996) found that low-income families (<$15,00/year) were 22-25 times more likely to suffer abuse or neglect. Wolfner and Gelles (1993) found that children of unemployed parents, or those working "blue collar jobs," were at increased risk for abuse. Other research has shown that children of parents who receive means-

tested support are at increased risk for abuse recidivism (Inkelas & Halfon, 1997; Levy, et al., 1995). Results from the present study are consistent with these findings. Within this large sample of Air Force families, there was a significant, but weak relationship between lower household income and abuse recurrence.

Length of Time Case Kept Open

Cases in which re-abuse occurred had, on average, had been monitored longer in response to the initial incident (8.6 months vs. 5.8 months). This finding contradicts a previous study that found that cases that were kept open longer were less likely to recidivate (Johnson & L'Esperance, 1984), but is somewhat consistent with findings from a study by Fluke and colleagues (1999) which found that cases that received post-investigative services were more likely to recidivate. A more consistent finding has been that the *intensity* of services reduces recidivism (Johnson, 1995; Lutrell, et al., 1995). A measure of intensity of treatment was not available to the present study.

One plausible interpretation of this finding in the present study is that caseworkers correctly identified higher risk families, deciding to monitor these cases longer. However, despite the increased attention, these families recidivated at higher rates than did others not chosen for further monitoring. Another plausible interpretation is that the increased surveillance afforded to these families resulted in identification of subsequent abusive incidents that would have gone undetected had the cases been closed earlier.

Referral Source

Of the six referral sources examined, only medical referrals demonstrated a significant correlation with re-abuse. Children who were referred to Family Advocacy by

a medical provider for the initial incident were 35% less likely to suffer a subsequent abuse as compared to those referred by other sources. Interestingly, Paddock (1995), analyzing a similar sample of U.S Air Force cases, found that referrals from medical providers were likely to involve *more severe injury* than were those from other sources. Other research has shown that referrals from anonymous reporters are less likely to be substantiated (Jason & Andereck, 1983; Zuravin, et al., 1987).

Victim/Offender Relationship

This study was limited to examining the victim/offender relationship in terms of whether the offender was a parent or non-parent, and thus is primarily an indicator of the offender's continued access to the victim. This variable demonstrated a bivariate correlation with re-abuse. Children experiencing recurrent abuse were 55% more likely to have suffered initial abuse from a parent than were children who suffered no recurrence of abuse. However, this relationship became non-significant in the prediction model when the all other predictors were entered as controls. There have been no previous reports in the literature regarding this variable.

Offender's Age

Like offender's relationship to the child, *offender's age* demonstrated a bivariate correlation with re-abuse, with children suffering re-abuse more likely to have been initially abused by a younger offender. However, the mean ages for the two groups differed by only one year (30.4 years vs. 31.3 years) and this relationship too became non-significant in the logistic regression prediction model. Previous research has reported

only that younger *mothers* are more likely to report *initial* abuse (Creighten, 1985; Zuravin, 1987).

Interactions

As noted, family income was negatively correlated with re-abuse. Additionally, income interacted with the offender's gender. Although higher income was associated with reduced risk for abuse for all groups, it was slightly more protective when the perpetrator was male. Findings regarding the *offender sex x income* interaction as somewhat consistent with Gelles' finding that poverty interacted with the gender of single parents, resulting in higher risk in mother-only families, but unrelated to violence in father-only families (1989). However, the present study found higher income to be associated with decreased violence when perpetrators were male or female, but slightly stronger when the abuser was male.

Non-significant Variables

The child's age, child's sex, offender's sex, offender's marital status, severity of initial abuse, substance involvement, and whether the offender received treatment were non-significant in bivariate and multivariate analyses. Also, four interaction terms: victim's race x income, offender's sex x substance involvement, child's sex x child's age, and child's age x child's age (curvilinear), were non-significant.

Previous research has shown the child's age to be associated with abuse. Fryer and Miyoshi (1994) found younger children more likely to be re-abused. Wolfner and Gelles (1993) found a curvilinear relationship for the child's age and abuse, with preschool aged children at higher risk than either toddlers or adolescents.

Child's gender has been reported to correlate with abuse in some studies, with girls more likely to be abused (Sedlack, 1997), and boys more likely to suffer severe injury (Rosenthal, 1988; Sedlack & Broadhurst, 1996). In the present study, the child's sex was not correlated with re-abuse, nor did it interact with the age of the child in relation to this outcome.

National incidence research on initial abuse has reported mothers to be overly represented as perpetrators in cases of neglect, and males overly represented as perpetrators of physical abuse (Sedlack & Broadhurst, 1996). The perpetrator's gender has not been reported as predictive of *re-abuse*. While offender's sex was non-significant in the present study, gender did interact with income, with higher income reducing risk in cases involving males and female offenders, however somewhat more so for male perpetrators.

Substance abuse has been consistently reported as a correlate of several outcomes related to child abuse. Parents who abuse substances have been found to be more likely to initially abuse and neglect their children (Chaffin, et al., 1996), and to re-abuse (English, et al., 2000). In the present study, information was limited to "substance involvement" that is, whether or not alcohol or other drug use was reported in related to the abuse incident. This is quite different than a history "substance abuse". Also, except in abuse cases identified immediately, field workers must rely on self-report or report of others present at the time of abuse to confirm substance involvement. Therefore, the reliability of these data is suspect. Ammerman (1999) found substance abuse interacted with gender, with substance abusing fathers more likely to report "abuse potential". The

interaction between offender sex and substance abuse was non-significant in relation to re-abuse in the present study.

The *treatment* variable in the present study was limited to a dichotomous measure; offender got treatment (yes/no). Lack of treatment provision may result when the Family Advocacy Case Management Team does not recommend services, or when a family is non-compliant with recommended treatment services. Previous research has shown the intensity of services to be negatively correlated with abuse recurrence (Johnson, 1995; Lutrell, et al., 1995). The inability of the present study to measure the amount of service contact reduced the likelihood of finding a significant relationship with recurrence.

Performance of Prediction Models

Camasso & Jagannathan (1995) conducted the only previously published study using ROC Curve analysis to evaluate risk assessment models. The areas under the ROC curve produced in the present study by the BLR (.65) and ANN models (.644) fell between those previously produced by the Washington Model (.68) and the Illinois Model (.58) in the previous study. This translates to approximately 60% correct classification when a diagnostic threshold that balances sensitivity and specificity evenly at about 60% is chosen, as opposed to increasing true positives or true negatives.

The goal of the present study to develop and test an artificial neural network model capable of predicting abuse recurrence was successful, inasmuch as the resultant model did predict significantly better than chance across a range of diagnostic thresholds.

However, Hypothesis 2, that the Bayesian ANN model would demonstrate better predictive validity than the BLR model was not confirmed. The BLR model actually produced predictions that were slightly, but significantly, better than those produced by the ANN. Further experimentation using a multi-layer perceptron, an alternative to the Bayesian network, produced predictive accuracy comparable to the regression model. In practical terms, all three models were equally accurate.

This finding is contrary to findings from numerous studies of similar prediction problems discussed above. But, as noted, neural networks have not outperformed other statistical methods in all comparisons. In fact, when compared to logistic regression, as opposed to linear regression or discriminant analysis, reported findings are more balanced.

For instance, Dwyer (1992) found that, although superior to discriminant analysis, a neural network was only on par with logistic regression in an application to predict bankruptcy. Ongphiphadhanakul and colleagues (1997) found no differences in terms of sensitivity or specificity between an ANN and logistic regression model for predicting low bone mineral density in postmenopausal women. Cloud (1999) found Radial Basis Function ANN and BLR model to be comparable in terms of predicting Alcoholics Anonymous affiliation from pretreatment factors.

In the present study, Hypothesis 2 (superiority of the ANN model) was based primarily upon the premise that unanticipated interactions and curvilinear relationships likely exist among the predictor variables, and that the ANN's ability to model these relationships without a priori specification would prove advantageous. However, it would

appear that the one statistically significant interaction derived from previous research findings produced an accurate representation of pertinent relationships among the predictors in the BLR model. Relationships between these variables appear to be more straightforward than anticipated. It should also be noted that the base rate of abuse recurrence (6%) within the present data set provided a particularly rigorous test of all of the models. As discussed in detail above, prediction of such rare events is exceedingly difficult.

Limitations

Population Differences

This study consisted of a retrospective, exploratory analysis of a large pre-existing electronic data file containing case characteristics of all reports of physical child abuse received by the U.S. Air Force Central Registry from 1990-2000. Generalization of findings from this study to other populations is inappropriate due to differences between this population and the civilian population. Military populations are likely to differ substantially from their civilian counterparts in terms of age, racial composition, income, and geographic mobility. Military families experience unique stressors, such as frequent relocation, separation from extended family, sudden deployments, among other factors. Conversely, they may be somewhat protected from other stressors that affect families receiving services from civilian CPS agencies, such as abject poverty, homelessness, and residence in very high crime neighborhoods.

Definitions and Base Rates

As discussed above, previous studies have defined re-abuse as either any subsequent referral (Johnson & L'Esperance, 1984; Marks & McDonald, 1989; Weedon, et al., 1988), or as only subsequent *substantiated* referrals (Baird, 1988, Johnson & Clancy, 1988). This distinction can dramatically alter the base rate of recurrence for a given population of interest. For example, using the former definition, Johnson and L'Esperance (1984) examined a sample with essentially equal percentages of recidivists and non-recidivists (N = 120). Whereas, data constraints limited the present study to only *substantiated* re-abuse incidents. Consequently, the base rate of recurrence in this study is only 6.25%. The present study was able to compensate for this problem to some extent due to the size of the overall sample (N = 5612). However, the small number of re-abuse examples proved problematic for adequate training of the Artificial Neural Network, which requires equivalent proportions of examples for each level of the target variable.

Differences in Observation Period

The present study was flawed in regard to variations in observation periods across cases. Although the method described above was adequate for identifying those families in which re-abuse occurred, it did not produce information regarding the time between the initial and subsequent incidents. Therefore, cases in which re-abuse occurred within a few months of the initial case closure were treated, statistically, the same as those in which the recurrence happened years after the first confirmed incident.

Also, initial abuse cases that were identified later in the observation period were observed for a far shorter time to determine if re-abuse occurred. Some of these families

may yet experience recurrent abuse, but were considered non-recidivists simply due to a

shortened period of observation. This may have resulted in a reduction in observed

between-groups differences and the limited the study's ability to detect true differences

between recidivist and non-recidivist cases.

As noted above, the ten-year observation period represents the high-end of those

used in previous research. Only Herrenkohl and colleagues (1979) allowed for a ten-year

observation period, whereas, other studies have limited the observation period to as little

as one year (Lutzker & Rice, 1987). The length of observation period affects observed

rates of recurrence (Fluke, et al., 1999). Additionally, the length of observation affects the

practical utility of the findings. Although factors found to be predictive of recurrence

within one year may drive case management and policy decisions, those associated with

abuse recurrence that occurs several years after the initial incident may have little

relevance to CPS policy and practice. The extraordinarily long observation period used in

the present study was chosen in order to identify as many recurrences as possible and

thus to optimize statistical power. This was accomplished at the cost of applicability of

the findings. The present study is exploratory in nature, aiming to increase general

knowledge of potential predictors of physical abuse recurrence and of statistical methods

with potential utility for assessing the predictive validity of risk assessment models.

Further research, preferable longitudinal in design, would be necessary in order to

confirm the usefulness of these variables in informing practice decisions.

<u>Unmeasured Variables</u>

The prediction models tested in the present study are likely misspecified due to

the absence of important predictors. Several variables that have demonstrated predictive

potential in previous research include: child's special needs (Burrell, et al., 1994;

Marshall & English, 2000; Weedon, et al., 1988), presence of domestic violence (English,

et al., 2000; DePanfilis & Zuravin, 1999; Ross, 1996), parental history of childhood

victimization (English, 1996; English & Pecora, 1994; Gil, 1970; Marshall & English,

2000; Shapiro, 1979), parenting skill (Johnson & L'Esperance, 1984), parent's

psychological well-being (Chaffin et al., 1996; DePanfilis & Zuravin, 1999), and family

size (Baird, 1988; DePanfilis & Zuravin, 1999; Johnson & L'Esperance, 1984; Wolfner

& Gelles, 1993; Zuravin, 1988).

It is likely that predictive accuracy would have been better in the present study

had measures of these variables been included. This weakness is typical of secondary

data analysis, especially when the data were not collected for the purpose of measuring

the particular research question.

Implications for Research, Policy, and Practice

<u>Research</u>

It is recommended that future studies attempting to predict recurrences of child

physical abuse, particularly those pertaining to military populations, examine the four

main effects variables and two interaction terms that demonstrated significance in the

present study. Also, previous research strongly supports the inclusion of parental

substance abuse in prediction models. The substance abuse measure used in this study was poorly operationalized. The present finding in regard to substance use should not be given undue weight.

The comparison of the BLR to the ANN models found the two to be generally comparable in terms of predictive accuracy across a broad range of diagnostic thresholds. However, the BLR model provides far more information to the researcher in regard to the contributions of individual variables. In fact, the ANN model was able to produce predictions better than chance only after variable elimination using BLR reduced the model to the five most pertinent variables.

Shadish (1986) advocates analyzing the same data using multiple data analytic strategies based on different assumptions in order to avoid reporting results that are methodologically biased. To this end, the two methods used in the current study (BLR and ANN) mutually support the conclusion that the included predictors can predict physical re-abuse within this population at approximately 60% accuracy, with comparable rates of sensitivity and specificity. Also, the ROC Curve analysis proved especially useful in examining the two models designed to predict a rare phenomenon, since data were provided across all classification thresholds.

Other than providing an alternative methodology by which to test the predictive power of these independent variables, the neural network offered no substantial advantage in the present study. The BLR model was superior both in terms of predictive power and in interpretability of results. However, as discussed above, the data in this study were derived from a military population, which differs substantially from that

encountered by civilian CPS agencies. Predictor variables may operate quite differently within these families as compared to civilian families. Likewise, the advantages of neural networks that have been reported in numerous prediction problems may materialize if applied to data from other populations.

Also, based on the substantial body of research indicating the superiority of artificial neural networks in prediction and classification problems characterized by numerous, undefined interactions among predictors, curvilinear relationships, and moderate levels of noise within the data; the potential of neural networks should be explored in relation to other clinical prediction problems. For instance, these models may prove useful for improving predictions of suicidal behavior, treatment compliance, and various treatment outcomes. Research should be directed toward evaluating artificial neural networks to make such clinical predictions.

Policy

Like many civilian CPS agencies, the Air Force Family Advocacy program directs that an intervention plan be implemented for substantiated cases (FAP Standards, 1998). However, research has raised questions regarding the usefulness of substantiation findings as a basis for intervention (DHHS, 1981 [NIS-1], 1996; Drake, 1996; Giovannoni, 1991; Jason, Andereck, Marks, & Tyler, 1982; Zuravin, et al., 1995). Substantiation decisions are dependent upon evidence or admission, and may not reflect ongoing risk or need for service intervention. A recent study found little difference in re-referral rates between cases that were initially founded and those that were unfounded (English et al., 2000).

The current Air Force policy of collecting a broad range of data only for substantiated cases precludes examination of re-referrals (unsubstantiated) and examination of re-abuse for cases initially substantiated versus those initially unsubstantiated. This policy also produces a low base-rate of observed recurrences, because only substantiated recurrences are identifiable. Predicting such rare events is especially difficult. In practical terms, Family Advocacy Officers may be interested in identifying *all* cases likely to be referred for subsequent allegation, even if those allegations are ultimately unsubstantiated due to lack of evidence.

The Air Force Family Advocacy Program collects data for a variety of uses. As stated above, the data set used in the present study was not collected for the purpose of building a risk assessment model. However, given the large number of abuse reports entered into the central registry each year, should measures of the missing variables discussed above be gathered, one could likely build a far more accurate model.

Family Advocacy clinicians currently assess risk by selecting a value from a Likert-type scale (1 – 4) that corresponds to a word-picture that best describes offender, child, and family characteristics. These ratings, which can be viewed as a kind of consensus-based risk assessment model, were not available to the present study. It would be useful to examine these the relationship between these ratings and abuse recurrence. If these ratings were included in the data collection, one could determine the predictive accuracy of these assessments, and if necessary, seek to improve predictions through empirical means. It may be that clinicians are already assessing risk of re-abuse accurately. However, this cannot be determined without these data.

Practice

Wald and Woolverton (1990) argue that for a predictive instrument to be used as a formal screening device in this arena it must have high specificity and extremely high sensitivity. The model tested in the present study does not meet this standard. In addition to limiting predictive validity, the absence of relevant variables resulted in a model, which although somewhat useful in predicting abuse recurrence, offers little in terms of informing practice decisions. The relevant predictors in the present study are those not amenable to clinical intervention. The present study is exploratory, and provides a base of knowledge regarding correlates of child physical abuse recurrence from data presently gathered by the central registry.

In order to produce a useful risk assessment tool, data on missing variables discussed above would have to be gathered. The resulting model would have to produce much better predictive accuracy to be useful in driving clinical decisions. Also, it would be inappropriate for caseworkers to base treatment decisions on demographic variables such as race and income, factors that may be markers for community and institutional biases in reporting and assessing abuse, rather than measures of actual risk.

Concluding Statement

Although neither research hypothesis was supported, the present study provides a starting point to understanding differences between families that return to Family Advocacy with confirmed incidents of physical child abuse and those who do not. The fact that the prediction models derived from binary logistic regression (BLR) and an

artificial neural network (ANN) were able to produce predictions significantly better than chance, using a very limited list of variables, suggests that these methods offer powerful tools for building and testing child abuse risk assessment models. Findings from this study favor the BLR method in terms of predictive accuracy and amount information provided to the researcher.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramczyk, L.W., & Swiegart, C. (1985). Child abuse and neglect: Indicated versus unfounded report characteristics. Final report to DHHS/OHDS/NCAAN for Grant Award #90-CA-0944.

Adams, W., Barone, N., & Tooman, P. (1982). The dilemma of anonymous reporting in child protective services. Child Welfare, 61, 3-14.

Air Force ServeNet, Family Advocacy Program. (2000). Available: http://airforcefap.org/fap/menu.asp

American Humane Association. (1988). Highlights of official child neglect and abuse reporting, 1986. Denver, CO: Author.

Ammerman, R.T., Kolko, D.J., Kirisci, L., Blackson, T.C., & Dawes, M.A. (1999). Child abuse potential in parents with histories of substance abuse disorder. Child Abuse & Neglect, 23(12), 1225-1238.

Anderson, R., Ambrosino, R., Valentine, D., & Lauderdale, M. (1983). Child deaths attributed to abuse and neglect: An empirical study. Children and Youth Services Review, 5, 75-89.

Ashizawa, K., MacMahon, H., Ishida, T., Nakamura, K., Vyborny, C.J., Katsuragawa, S., & Doi, K. (1999). Effect of an artificial neural network on radiologists' performance in the differential diagnosis of interstitial lung disease using chest radiographs. American Journal of Roentgenology, 172(5), 1311-5.

Baird, C. (1988). Development of risk assessment indices for the Alaska Department of Health and Social Services. In T. Tatara (Ed), <u>Validation research in CPS risk assessment: Three recent studies</u> (pp. 85-121). Washington, DC: American Public Welfare Association.

Baird, C., Ereth, J., & Wagner, D. (1999). <u>Research-based risk assessment: Adding equity to CPS decision-making</u> [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Baird, C., Wagner, D., Caskey, R., & Neuenfeldt, D. (1995). <u>The Michigan Department of Social Services Structured Decision Making System: An evaluation of its impact on child protection services</u> [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Baird, C., Wagner, D., Healy, & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. <u>Child Welfare, 78</u>(6), 723-748.

Baumann, D.J., Esterline, J.A., Zuniga, G., Smith, S., Whiteside, D., Fluke, J., Goertz, B., & Cohen, M. (1997). The implementation of risk assessment. In <u>Worker Improvement the the Structured Decision and Outcome Model.</u> Texas Department of Protective and Regulatory Services.

Baxt, W.G., & Skora, J. (1996). Prospective validation of artificial neural network trained to identify myocardial infarction. <u>The Lancet, 347,</u> 12-15.

Bergman, A.B., Larsen, R.M., & Mueller, B.A. (1986). Changing spectrum of serious child abuse. <u>Pediatrics, 77,</u> 113-116.

Berkowitz, S. (1991). Key findings on definitions of risk assessment to children and uses of risk assessment by state CPS agencies from the state survey component of the Study of High Risk Child Abuse and Neglect groups. Paper presented at the National Center on Child Abuse and Neglect. Washington, DC.

Bolen, R.M. (1998). Predicting risk to be sexually abused: A comparison of logistic regression to event history analysis. Child Maltreatment, 3(2), 157-170.

Brissett-Chapman, S. (1997). Child protection risk assessment and African American children: Cultural ramifications for families and communities. Child Welfare, 74(1), 45-63.

Brodzinski, J.D., Crable, E.A., & Sherer, R.F. (1994). Using artificial intelligence to model juvenile recidivism patterns. Computers in Human Services, 10(4), 1-19.

Bryce, T.J., Dewhirst, M.W., Floyd, C.E., Hars, V., & Brizel, D.M. (1998). Artificial neural network model of survival in patients treated with irradiation with and without concurrent chemotherapy for advanced carcinoma of the head and neck. International Journal of Radiation, Oncology, Biology, and Physics, 41(2), 339-45.

Buchanan, B.G., & Shortliffe, E.H. (Eds.). (1984). Rule-based expert systems. Reading, MA:Addison-Wesley.

Burrell, B., Thompson, B., & Sexton, D. (1994). Predicting child abuse potential across family types. Child Abuse & Neglect, 18(12), 1039-1049.

Calica, R., Cotton, E., & Edwards, M. (1998). Illinois child endangerment risk assessment protocol: Impact on short-term recurrence rates-year two. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Camasso, M.J., & Jagannathan, R. (1995). Prediction accuracy of the Washington and Illinois risk assessment instruments: An application of receiver operating characteristic curve analysis. Social Work Research, 19(3), 174-183.

Campana, A., Duci, A., Gambini, O., & Scarone, S. (1999). An artificial neural network that uses eye-tracking performance to identify patients with schizophrenia. Schizophrenia Bulletin, 25(4), 789-799.

Chaffin, M., Kelleher, & Hollenberg, J. (1996). Onset of physical abuse and neglect: Psychiatric, substance abuse, and social risk factors from prospective community data. Child Abuse & Neglect, 3, 191-203.

Child Abuse Prevention and Treatment Act (CAPTA) (P.L. 93-247, Jan. 1996 version), 42 U.S.C. 5101. Originally passed in 1974.

Church, K.B., & Curram, S.P. (1996). Forecasting consumers' expenditure: A comparison between econometric and neural network models. International Journal of Forecasting, 12(2), 255-67.

Cohen, J. (1977). Statistical power analyses for the behavioral sciences. New York: Academic Press.

Collier, A.F., McClure, F.H., Collier, J., Otto, C., & Polloi, A. (1999). Culture-Specific views of child maltreatment and parenting styles in a Pacific-Island community. Child Abuse & Neglect, 23(3), 229-244.

Collins, J.M., & Clark, M.R. (1993). An application of the theory of neural computation to the prediction of workplace behavior: an illustration and assessment of network analysis. Personal Psychology, 46, 503-524.

Connelly, C.D., & Straus, M.A. (1992). Mother's age and risk for physical abuse. Child Abuse & Neglect, 16, 709-718.

Cooil, B., Winer, R.S., & Rados, D.L. (1987). Cross-validation for prediction. Journal of Marketing Research, 24, 271-279.

Creighton, S.J. (1985). Am epidemiological study of abused children and their families in the United Kingdom between 1977 and 1982. Child Abuse and Neglect, 9, 441-448.

Cross, S.S., Harrison, R.F., & Kennedy, R.L. (1995). Introduction to neural networks. The Lancet, 346, October 21, 1075-1079.

Daley, M.R., & Pilavin, I. (1982). "Violence against children" revisited: Some necessary clarification of findings from a major national study. Journal of Social Service Research, 5(1/2), 61-81.

Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgement. Science, 243, 1668-1674.

DePanfilis, D., & Scannapieco, M. (1994). Assessing the safety of children at risk of maltreatment: Decision-making models. Child Welfare, 73(3), 229-245.

DePanfilis, D., & Zuravin, S.J. (1999). Epidemiology of child maltreatment recurrences. Social Service Review, 73(2), 218-239.

DePanfilis, D., & Zuravin, S.J. (1999). Predicting child maltreatment recurrences during treatment. Child Abuse & Neglect, 23(8), 729-743.

Department of Defense. (1986). Family Advocacy Program (Directive 6400.1) Washington, D.C.: Author.

Department of Defense. (1987). Child and Spouse Abuse Report (Directive 6400.2). Washington, D.C.: Author.

Department of Defense. (1992). Family Advocacy Program (Directive 6400.1) Washington, D.C.: Author.

Department of Health and Human Services, National Center on Child Abuse and Neglect. (1995). Child maltreatment 1993: Reports from the states to the Center on Child Abuse and Neglect. Washington, DC: U.S. Government Printing Office.

de Tommaso, M., Sciruicchio, V., Bellotti, R., Castellano, M., Tota, P., Guido, M., Sasanelli, G., & Puca, F. (1997). Discrimination between migraine patients and normal subjects based on steady state visual evoked potentials: Discriminant analysis and artificial neural network classifiers. Functional Neurology, 12(6), 333-338.

Doueck, H.J., English, D.J., DePanfilis, D., & Moote, G.T., jr. (1993). Decision making in child protective services: A comparison of selected risk assessment systems. Child Welfare, 72, 441-453.

Doueck, H.J., Levine, M., & Bronson, D.E. (1993). Risk assessment in child protective services: An evaluation of the Child at Risk Field System. Journal of Interpersonal Violence, 8(4), 446-467.

Doueck, H.J., & Lyons, P. (1998). The Child Well-Being Scales as a predictor of caseworker attention and services in child protection: A preliminary analysis. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Downing, J.D., Wells, S.J., & Fluke, J. (1990). Gatekeeping in child protective services: A survey of screening policies. Child Welfare, 69(4), 357-369.

Drake, B. (1996). Unraveling "unsubstantiated". Child Maltreatment, 1(5), 261-271.

Dwyer, M. (1992). A comparison of statistical techniques and artificial neural network models in corporate bankruptcy prediction. Doctoral dissertation, University of Wisconcin-Madison.

Dybowski, R., Weller, P., Chang, R., & Grant, V. (1996). Prediction of outcome in critically ill patients using artificial neural network synthesized by genetic algorithm. The Lancet, 347(9009), 1146-51.

El-Solh, A.A., Hsiao, C., Goodnugh, S., Serghani, J., & Grant, B.J.B. (1999). Predicting active pulmonary tuberculosis using an artificial neural network. Chest, 116(4), 968-973.

English, D.J. (1996). The promise and reality of risk assessment. Protecting Childern, 12(2), 9-13.

English, D.J., Marshall, D.B., Coghlan, L., Brummel, S., & Orme, M. (2000). Causes and consequences of the substantiation decision in Washington state child protective services. Children & Youth Services Review, to appear.

English, D.J., & Pecora, P.J. (1994). Risk assessment as a practice method in child protective services. Child Welfare, 73(5), 451-468.

Everson, H.T., er al. (1994). Using artificial neural networks in educational research: some comparisons with linear statistical models. Paper presented a the Annual Meeting of the Council on Measurement in Education, New Orleans, LA, 5-7 April.

Fanshel, D., Finch, S.J., & Grundy, J.F. (1994). Testing the measurement properties of risk assessment instruments in child protective services. Child Abuse & Neglect, 18, 1073-1084.

Farmer, R.M., Medearis, A.L., Hirata, G.I., & Platt, L.D. (1992). The use of a neural network for the ultrasonographic estimation of fetal weight in the macroscopic fetus. American Journal of Obstetrics and Gynecology, 166(5), 1467-1472.

Ferrara, J.M., Parry, J.D., & Lubke, M.M. (1985). CLASS.1: An expert system for student classification. Technical paper (ERIC Document Reproduction Service No. ED 263-734), Utah State University, Logan, UT.

Flango, V.E. (1991). Can central registries improve substantiation rates in child abuse and neglect cases? Child Abuse & Neglect, 15, 403-413.

Fluke, J. (1993). The use of expert systems and neural networks in CPS risk assessment: What are the possibilities and limitations. Summary of small group discussion at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J. (1994). Emerging critical conceptual issues in risk assessment research and practice. Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J., Wells, S., England, P., & Gamble, T. (1993). Evaluation of the Pennsylvania approach to risk assessment. Paper presented at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Fluke, J., Edwards, M., & Wells, S. (1996). Evaluation of the Illinois child endangerment risk assessment protocol: Safety domain. Unpublished manuscript.

Fluke, J.D., Yuan, Y.T., & Edwards, M. (1999). Recurrence of maltreatment: An application of the National Child Abuse and Neglect Data System (NCANDS). Child Abuse & Neglect, 23(7), 633-650.

Fryer, G.E., & Miyoshi (1994). A survival analysis of the revictimization of children: The case of Colorado. Child Abuse & Neglect, 18(12), 1063-1071.

Fuller, T.L., & Wells, S.J. (1998). Illinois child endangerment risk assessment protocol: A case control study of short-term recurrence among intact family cases. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Galletly, C.A., Clark, C.R., & McFarlane, A.C. (1996). Artificial neural networks: A prospective tool for the analysis of psychiatric disorders. Journal of Psychiatry and Neuroscience, 21(4), 239-247.

Gallinari, P., Thiria, S., Badran, F., & Fogelman-Soulie, F. (1991). On the relations between discriminant analysis and multilayer perceptrons. Neural Networks, 4, 349-360.

Garson, G.D. (1998). Neural Networks: An Introductory Guide for Social Scientists. Thousand Oaks, California. Sage.

Gaudin, J.M., Jr., Polansky, N., & Kilpatrick, A.C. (1992). The Child Well-Being Scales: A field trial. Child Welfare, 17, 319-328.

Gelles, R.J. (1982). Child abuse and family violence: Implications for medical professionals. In Newberger, E.H. (Ed), Child Abuse. Boston. Little Brown.

Gelles, R.J. (1989). Child abuse and violence in single-parent families: Parent absence and economic deprivation. American Journal of Orthopsychiatry, 59(4), 492-501.

Gil, D.G. (1970). Violence against children. Cambridge, MA: Harvard University Press.

Gingerich, W.J. (1990). Expert systems and their potential uses in social work. Families in Society, 71(4), 220-228.

Giovannoni, J.M., & Becerra, R.M. (1979). Defining child abuse. New York. Free Press.

Goodman, H., Gingerich, W.J., & de Shazer, S. (1989). BRIEFER: An expert system for clinical practice. Computers in Human Services, 5(1/2), 53-68.

Gordon, J.S. (1991). (Probability correct classification as a function of increasing decision thresholds). Unpublished raw data.

Gordon, J.S. (1992). A neural network approach to the prediction of violence. Unpublished doctoral dissertation. Oklahoma State University.

Greenland, C. (1987). Preventing CAN deaths: An international study of deaths due to child abuse and neglect. London: Tavistock Publications.

Hampton, R.L. (1987). Race, class, and child maltreatment. Journal of Comparative Family Studies, 18, 113-126.

Hanley, J.A., & McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143, 29-36.

Hanley, J.A., & McNeil, B.J. (1983). A method of comparing the areas under receiver operating curves derived from the same cases. Radiology, 148, 839-843.

Hartzberg, J., Stanley, J., & Lawrence, M. (1990). Brainmaker user's guide and reference manual (Computer program manual). Sierra Madre, CA: California Scientific Software.

Heckert, T.M. (1994). Effect of base rate on the applicability of neural network analysis for classification. Doctoral dissertation, Bowling Green State University.

Hedlund, J.L., Viewing, B.W., & Cho, D.W. (1987). Computer consultation for emotional crises: An expert system for "non-experts." Computers in Human Behavior, 3, 109-127.

Herrenkohl, R.C., Herrenkohl, E.C., Egolf, B., & Seech, M. (1979). The repetition of child abuse: How frequently does it occur? Child Abuse & Neglect, 3, 67-72.

Hiemstra, Y. (1996). Linear regression versus backpropagation networks to predict quarterly stock market excess returns. Computational Economics, 9(1), 67-76.

Holder, W., & Corey, M.K. (1993). The Child At Risk Field System: Forms and instructions booklet. Charlotte, NC: Action for Child Protection.

Howing, P.T., Wodarski, J.S., Kurtz, P.D., & Gaudin, J.M., Jr. (1989).

Methodological issues in child maltreatment research. Social Work Research and

Abstracts, 25(3), 3-7.

Hutchinson, E.D. (1989). Child protective screening decisions: An analysis of

predictive factors. Social Work Research and Abstracts, 25(3), 9-15.

Inkelas, M, & Halfon, N. (1997). Recidivism in child protective services.

Children and Youth Services Review, 19, 139-161.

Jackson, A. (1983). Professional and non-professional sources of child

maltreatment reports. For Your Information, 4, 3-7.

Jason, J., & Andereck, N.D. (1983). Fatal child abuse in Georgia: The

epidemiology of severe physical child abuse. Child Abuse & Neglect, 7, 1-9.

Jefferson, M.F., Pendleton, N., Lucas, C.P., Lucas, S.B., & Horan, M.A. (1998).

Evolution of artificial neural network architecture: Prediction of depression after mania.

Methods of Information in Medicane, 37, 220-225.

Johnson, W. (1993). Maltreatment recurrence as a criterion for validating risk

assessment instruments. Paper presented at the Seventh National Roundtable on CPS

Risk Assessment. San Francisco, California.

Johnson, W. (1994). Culturally sensitive risk assessment: A conceptual

framework and supporting data. Paper presented at the Eighth National Roundtable on

CPS Risk Assessment. San Francisco, California.

Johnson, W. (1996). Risk assessment research: progress and future directions.

Protecting Children, 12(2), 14-19.

Johnson, W., & Clancy, T. (1988). A study to find improved methods of screening and disposing of reports of child maltreatment in the emergency response program in Alameda County, California. In Tatara (Ed), Validation research in CPS risk assessment: Three recent studies. Occasional Monograph Series, No. 2 (pp47-84). Washington, DC: American Public Welfare Association.

Johnson, W., & Clancy, T. (1989). Preliminary findings from a study of risk assessment accuracy and service effectiveness in home-based services physical abuse cases. In T. Tatara (Ed.), Third National Roundtable on CPS Risk Assessment: Summary of highlights. Washington, DC: American Public Welfare Association.

Johnson, W., & L'Esperance, J. (1984). Predicting the recurrence of child abuse. Social Work Research and Abstracts, 20(2), 21-26.

Kennedy, R.L., Harrison, R.F., Burton, A.M., Fraser, H.S., Hamer, W.G., MacArthur, D., McAllum, R., Steedman, D.J. (1997). An artificial neural network system for diagnosis of acute myocardial infarction (AMI) in the accident and emergency department: evaluation and comparison with serum myoglobin measurements. Computer Methods and Programs in Biomedicine, 52, 93-103.

Kolko, D.J., (1998). CPS operations and risk assessment in child abuse cases receiving services: Initial findings from the Pittsburgh service delivery study. Child Maltreatment, 3(3), 262-275.

Korbin, J.E. (1981). Very few cases of child abuse and neglect in the People's Republic of China. In J.E. Korbin (Ed.), Child abuse: Cross-cultural perspectives. Berkeley, CA: University of California Press.

Kruttschnitt, C., Mcleod, J.D., & Dornfeld, M. (1994). The economic environment of child abuse. Social Problems, 41(2), 299-315.

Law, J.R., Maumann, D.J., Gober, K.J., Schultz, D.F., Ohmart, R., & Kern, H. (1997). Phase III: Evaluating the effectiveness of an actuarial risk assessment model. In Worker Improvement the the Structured Decision and Outcome Model. Texas Department of Protective and Regulatory Services.

Law, J.R., Kern, H.D., Schultz, D.F., Gober, K.J., Schwab, J., & Baumann, D.J. (1997). Screening models for intake and investigation. In Worker Improvement the the Structured Decision and Outcome Model. Texas Department of Protective and Regulatory Services.

Law, R., McFadden, T., & Kern, H. (1994). Risk assessment at intake and at investigation. Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Lawrence, J. (1993). Introduction to neural networks: Design, theory and application (5th ed.) Nevada City, CA: California Scientific Software.

Levy, H.B., Markovic, J., Chaudhry, U., Ahart, S., & Torres, H. (1995). Reabuse rates in a sample of children followed for 5 years after discharge from a child abuse inpatient assessment program. Child Abuse & Neglect, 19, 1363-1377.

Lindsey, D. (1991). Factors affecting the foster care placement decision: Analysis of national survey data. American Journal of Orthopsychiatry, 61, 272-281.

Lindsey, D. (1992). Reliability of the foster care placement decision: A review. Research on Social Work Practice, 2, 65-80.

Lippman, R.P. (1987). An introduction to computing with neural nets. IEEE ASSP Magazine, April, pp. 4-22.

Luttrell, J., Hull, S., & Wagner, D. (1995). The Michigan Department of Social Services structured decision-making system. Paper presented as the Ninth National Roundtable on CPS Risk Assessment. San Francisco, California.

Lutzker, J.R., & Rice, J.M. (1987). Using recidivism data to evaluate Project 12-Ways: An ecobehavioral approach to the treatment and prevention of child abuse and neglect. Journal of Family Violence, 2, 283-290.

Lyons, P., Doueck, H.J., Koster, A.J., Witzky, M.K., & Kelly, P.L. (1999). The Child Well-Being Scales as a clinical tool and a management information system. Child Welfare, 78(2), 241-258.

Lyons, P., Doueck, H.J., & Wodarski, J.S. (1996). Risk assessment for child protective services: A review of the empirical literature on instrument performance. Social Work Research, 20(3), 143-155.

Magura, S., Moses, B.S., & Jones, M.A. (1987). Assessing risk and measuring change in families: The Family Risk Scales. Washington, DC: Child Welfare League of America.

March, J.S., & Curry, J.F. (1998). Predicting the outcome of treatment. Journal of Abnormal Child Psychology, 26(1), 39-51.

Marks, J., & McDonald, T.P. (1989). Risk assessment in child protective services: Predicting recurrence of child maltreatment. Portland, ME: National Child Welfare Resource Center.

Marshall, D.B., & English, D.J. (2000). Neural network modeling of risk assessment in child protective services. Psychological Methods, 5(1) in press.

Martindale, E.S., Ferrara, J.M., & Campbell, B.W. (1987). A preliminary report on the performance of CLASS.LD. Computers in Human Behavior, 3, 263-272.

McDonald, T., & Marks, J. (1991). A review of risk factors assessed in child protective services. Social Service Review, 65(1), 112-132.

Meehl, P.E. (1954). Clinical versus statistical prediction. Minneapolis, Minnesota. University of Minnesota Press.

Meyer, B.L. (1998). Implementing actuarial risk assessment: Policy decisions and field practice in New Mexico. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Meyer, B.L., & Wagner, D. (1998). Using actuarial risk assessment to identify unsubstantiated cases for preventative intervention in New Mexico. Paper presented at the Twelfth National Roundtable on CPS Risk Assessment. San Francisco, California.

Milner, J.S., & Wimberley, R.C. (1979). An inventory for the identification of child abusers. Journal of Clinical Psychology, 35, 95-100.

Milner, J.S. (1995). Physical child abuse assessment: Perpetrator evaluation. In J.C. Campbell (Ed.), Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 41-67). Thousand Oaks, California. Sage.

Milner, J.S., & Campbell, J.C. (1995). Prediction issues for practitioners. In J.C. Campbell (Ed.), Assessing dangerousness: Violence by sexual offenders, batterers, and child abusers (pp. 20-40). Thousand Oaks, California. Sage.

Mollerstrom, W.W., Patchner, M.A., & Milner, J.S. (1995). Child maltreatment: the United States Air Force's response. Child Abuse & Neglect, 19(3), 325-

Mulsant, B., & Servan-Schreiber, D. (1984). Knowledge engineering: A daily activity on a hospital ward. Computers in Biomedical Research, 17, 71-91.

Munro, E. (1999). Common errors of reasoning in child protection work. Child Abuse & Neglect, 23(8). 745-758.

Murphy-Berman, V. (1994). A conceptual framework for thinking about risk assessment and case management in child protective service. Child Abuse & Neglect, 18(2), 193-201.

National Center for Child Abuse and Neglect (1981). National study of the incidence and severity of child abuse and neglect. 81-30325. Washington, DC: U.S. Department of Health and Human Services. [NIS-1].

National Center for Child Abuse and Neglect (1988). National study of the incidence and severity of child abuse and neglect. Washington, DC: U.S. Department of Health and Human Services. [NIS- 2].

National Center for Child Abuse and Neglect (1996). National study of the incidence and severity of child abuse and neglect. 105-91-1800. Washington, DC: U.S. Department of Health and Human Services. [NIS- 3].

National Council on Crime and Delinquency (NCCD), (2000). Child abuse and neglect: Improving consistency in decision-making [On-line]. Available: http://www.nccd-crc.org/crcindex.htm

Nasuti, J.P., & Pecora, P.J. (1993). Risk assessment scales in child protection: A test of the internal consistency and interrater reliability of one statewide system. Social Work Research and Abstracts, 29(2), 28-33.

Neuenfeldt, D. & DeMares, M. (1994). Risk assessment validation study and its use in four urban Wisconsin counties. Paper presented at the Eighth National Roundtable on CPS Risk Assessment. San Francisco, California.

Norusis, M.J. (1997). SPSS Guide to Data Analysis. Upper Saddle River, NJ: Prentice-Hall, Inc.

Nunnally, J.C. (1967). Psychometric Theory. New York, New York. McGraw-Hill.

Ongphiphadhanakul, B., Rajatanavin, R., Chailurkit, L., Piaseu, N., Teerarungsikul, K., Sirisriro, R., Komindr, S., & Pauvilai, G. (1997). Prediction of low bone mineral density in post menopausal women by artificial neural network model compared to logistic regression model. Journal of the Medical Association of Thailand, 80(8), 508-515.

Orme, J.G. (1986). Erroneous use of observed percentage agreement in discriminant function analysis. Social Work Research and Abstracts, 22, 2.

Paddock, J.B. (1995). Factors predictive of injury severity in cases of physical child abuse among Air Force families: A cross-validation study. Unpublished dissertation.

Parton, N., Thorpe, D., & Wattam, C. (1997). Child Protection: Risk and the Moral Order. London. Macmillan Press.

Patterson, D.A. & Cloud, R.N. (2000). The application of artificial neural networks for outcome prediction in a cohort of severely mentally ill outpatients. Forthcoming in The Journal of Technology for Human Services.

Pecora, P.J. (1991). Investigating allegations of child maltreatment: The strengths and limitations of current risk assessment systems. Child and Youth Services, 15(2), 73-92.

Picard, R.R., & Cook, R.D. (1984). Cross-validation of regression models. Journal of the American Statistical Association, 79(387), 575-583.

Pofahl, W.E., Walczak, S.M., Rhone, E., & Izenberg, S. (1998). Use of an artificial neural network to predict length of stay in acute pancreatitis. The American Surgeon, 64(9), 868-872.

Pugh, G.A. (1991). A comparison of neural networks to SPC charts. Computers and Industrial Engineering, 21(1), 253-55.

Ross, S.M. (1996). Risk of physical abuse to children of spouse abusing parents. Child Abuse & Neglect, 20(7), 589-598.

Rosenthal, J.A. (1988). Patterns of reported child abuse and neglect. Child Abuse & Neglect, 12, 263-271.

Ross, S.M. (1996). Risk of physical abuse to children of spouse abusing parents. Child Abuse & Neglect, 20(7), 589-598.

Rossi, P., Schuerman, & Budde, S. (1996). Understanding child maltreatment decisions and those who make them. Chicago: University of Chicago, Chapin Hall Center for Children.

Rubin, A., & Babbie, E. (1997). Research methods for social work (3rd Ed.), Pacific Grove, CA: Brooks/Cole.

Ruscio, J. (1998). Information integration in child welfare cases: An introduction to statistical decision-making. Child Maltreatment, 3(2), 143-156.

Salovitz, B. (1992). New York State risk assessment and service planning model: A review of the developmental process and key features of the model. Paper presented at the Sixth National Roundtable on CPS Risk Assessment. San Francisco, California. American Public Welfare Association.

Schene, P. (1996). The risk assessment roundtables: a ten-year perspective. Protecting Children, 12(2), 4-8.

Schoech, D. (1999). Human service technology: Understanding, designing, and implementing computer and internet applications in the social services. New York, New York. The Haworth Press, Inc.

Schuerman, J.R. (1987). Expert consulting systems in social welfare. Social Work Research and Abstracts, 23(3), 14-18.

Schuerman, J.R., & Vogel, L.H. (1986). Computer support of placement planning: The use of expert systems in child welfare. Child Welfare, 65(6), 531-543.

Seaberg, J.R. (1977). Predictors of injury in physical child abuse. Journal of Social Service Research, 1(1), 63-76.

Sedlak, A.J. (1997). Risk factors for the occurrence of child abuse and neglect. Journal of Aggression, Maltreatment, and Trauma, 1(1), 149-187.

Sedlack, A.J., & Broadhurst, D.B. (1996). Executive summary of the third national incidence study of child abuse and neglect. U.S. Department of Health and Human Services 105-91-1800.

Serna, R.W., Baer, R.D., & Ferrara, J.M. (1986). An expert system for behavior analysis. Paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, WI.

Shadish, W. H., Jr. (1986). Planned critical multiplism: Some elaborations. Behavioral Assessment, 8, 75-103.

Shapiro, D. (1979). Parents and protectors: A study of child abuse and neglect. New York Research Center, Child Welfare League of America.

Sheets, D.A. (1996). Caseworkers, computers and risk assessment: A promising partnership. APSAC Advisor, 9(1), 7-12.

Sohl, J.-H.E. & Venkatachalalm, A.R. (1995). A neural network approach to forcasting-model selection, Information Management, 29(6), 297-303.

SPSS 10.0 (1999). [Computer program]. Chicago: SPSS Inc.

SPSS (1999) SPSS 10.0 User's Guide, Chicago: Author.

SPSS Neural Connection 2.0 (1997). [Computer program]. Chicago: SPSS Inc.

SPSS (1997). Neural Connection 2.0 User's Guide, Chicago: Author.

Squadrito, E. & Wagner, D. (1993). Family risk assessment study by the Rhode Island Department of Children, Youth, and Families and the National Council on Crime and Delinquency/Children's Research Center. Paper presented at the Seventh National Roundtable on CPS Risk Assessment. San Francisco, California.

Starr, R.H. (Ed.). (1982). Child abuse prediction: Policy implications. Cambridge, Massachusetts: Ballinger.

Tailor, A., Jurkovic, D., Bourne, T.H., Collins, W.P., & Campbell, S. (1999). Sonographic prediction of malignancy in adnexal masses using an artificial neural network. British Journal of Obstetrics and Gynaecology, 106, 21-30.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62(2), 157-176.

Timmerman, D., Verrelst, H., Bourne, T.H., De Moor, B., Collins, W.P., Vergote, I., & Vandewalle, J. (1999). Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound in Obstetrics and Gynecology, 13(1), 17-25.

U.S. Department of Health and Human Services. Child maltreatment (1997): Reports from the states to the National Child Abuse and Neglect Data System (Washington, DC: U.S. Government Printing Office.

Valentine, D.P., Acuff, D.S., Freeman, M.L., & Andreas, T. (1984). Defining child maltreatment: A multidisciplinary overview. Child Welfare, 63(6), 497-509.

Vogt, W.P. (1999). Dictionary of statistics and methodology: A nontechnical guide for the social sciences. Thousand Oaks, CA: Sage.

Wald, M.S., & Woolverton, M. (1990). Risk assessment: "The emperor's new clothes?" Child Welfare, 69, 483-511.

Weedon, J., Torti, T.W., & Zunder, P. (1988). Vermont Division of Social Services Family Risk Assessment Matrix: Research and Evaluation. In Tatara (Ed), Validation research in CPS risk assessment: Three recent studies (pp. 3-43). Washington, D.C.: American Public Welfare Association.

Weiss, S.M., & Kurlikowski, C.A. (1991). Computer systems that learn: Classification and prediction methods from statistics, neural nets, machine learning, and expert systems. San Mateo, CA: Morgan Kaufmann.

Wells, S.J., & Anderson, T. (1992). Workers' estimation of risk as a predictor of case substantiation. Washington, DC: American Public Welfare Association.

Wells, S.J., & Anderson, T. (1992). Model building in child protective services intake and investigation. Final report to the National Center on Child Abuse and Neglect for Grant #90-CA-1407. Washington, DC: American Bar Association Center on Children and the Law.

White, H. (1989). Some asymptotic results for learning in single hidden-layer feedforward neural models. Journal of the American Statistical Society, 84, 1003-1013.

Wolfner, G.D., & Gelles, R.J. (1993). A profile of violence toward children: A national study. Child Abuse & Neglect, 17, 197-212.

Zaknich, A. (1998). Artificial neural networks: An introductory course [On-line]. Available: http://www.maths.uwa.edu.au/~rkcalley/ann_all/ann_all.html

Zernikow, B., Holtmannspoetter, K., Michel, E., Theilhaber, M., Pielemeier, W., & HenneckeK.H. (1998). Artificial neural network for predicting intercranial haemorrhage in preterm neonates. Acta Paediatrica, 87(9), 969-75.

Zernikow, B., Holtmannspoetter, K., Michel, E., Pielemeier, W., Hornschuh, F., Westermann, A., & Hennecke, K.H. (1998). Artificial neural network for risk assessment in preterm neonates. Archives of Disease in Childhood, 79(2), F129-34.

Zhang, Z., Barnhill, S.D., Zhang, H., Xu, F., Yu, Y., Jacobs, I., Woolas, R.P., Berchuck, A., Madyastha, K.R., & Bast, R.C, jr. (1999). Combination of multiple serum markers using an artificuial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecologic Oncology, 73, 56-61.

Zhao, R., Xu, G., Yue, B., Liebich, H.M., & Zhang, Y. (1998). Artificial neural network classification based on capillary electrophoresis of urinary nucleosides for the clinical diagnosis of tumors. Journal of Chromatography A, 828, 489-496.

Zou, Y., Shen, Y., Shu, L., Wang, Y., Feng, F., Xu, K., Qu, Y., Song, Y., Zhong, Y., Wang, M., & Liu, W. (1996). Artificial neural network to assist psychiatric diagnosis. British Journal of Psychiatry, 169(1), 64-67.

Zuravin, S. (1988). Child maltreatment and teenage first birth: A relationship mediated by chronic sociodemographic stress? American Journal of Orthopsychiatry, 58, 91-103.

Zuravin, S.J., Orme, J.G., & Hegar, R.L. (1991). Factors predicting substantiation and severity of child abuse injury. In National Center on Child Abuse and Neglect: Symposium on Risk Assessment in Child Protective Services (Vol. 4). Washington, DC: National Center on Child Abuse and Neglect.

Zuravin, S.J., Orme, J.G., Hegar, R.L. (1994). Predicting severity of child abuse injury with ordinal probit regression. Social Work Research, 18(3), 131-138.

Zuravin, S.J., Orme, J.G., Hegar, R.L. (1995). Disposition of child physical abuse reports: Review of the literature and test of a predictive model. <u>Children and Youth Services Review, 17</u>(4), 547-566.

Zuravin, S.J., Watson, B., & Ehrenschaft, M. (1987). Anonymous reports of child physical abuse: Are they as serious as reports from other sources? <u>Child Abuse & Neglect, 11,</u> 521-529.

VITA

Captain Christopher W. Flaherty was born in Paducah, Kentucky on March 25, 1965. He graduated from Crittenden County High School in 1983, and enlisted in the U.S. Air Force in 1985, where he worked as a psychiatric technician. He was awarded a Bachelor of Social Work from the University of Southern Colorado in 1990 and a Master of Social Work from the University of Denver in 1992.

He was commissioned as an officer in the United States Air Force Biomedical Science Corps in December of 1992, and spent the next five years working as a clinical social worker at bases in South Dakota and California. In the fall of 1998, the author began Ph.D. studies at the University of Tennessee, Knoxville, in the college of Social Work under sponsorship by the Air Force Institute of Technology.

Captain Flaherty married his wife, Johnina in 1991. They have two children, Elizabeth and Patrick.